

1-1-2011

## Cutting through the clutter: Searching for targets in evolving complex scenes

Mark B. Neider  
*University of Central Florida*

Gregory J. Zelinsky

Find similar works at: <https://stars.library.ucf.edu/facultybib2010>  
University of Central Florida Libraries <http://library.ucf.edu>

This Article is brought to you for free and open access by the Faculty Bibliography at STARS. It has been accepted for inclusion in Faculty Bibliography 2010s by an authorized administrator of STARS. For more information, please contact [STARS@ucf.edu](mailto:STARS@ucf.edu).

---

### Recommended Citation

Neider, Mark B. and Zelinsky, Gregory J., "Cutting through the clutter: Searching for targets in evolving complex scenes" (2011). *Faculty Bibliography 2010s*. 1718.  
<https://stars.library.ucf.edu/facultybib2010/1718>

# Cutting through the clutter: Searching for targets in evolving complex scenes

**Mark B. Neider**

Department of Psychology, University of Central Florida,  
Orlando, FL, USA



**Gregory J. Zelinsky**

Department of Psychology, Stony Brook University, USA



We evaluated the use of visual clutter as a surrogate measure of set size effects in visual search by comparing the effects of subjective clutter (determined by independent raters) and objective clutter (as quantified by edge count and feature congestion) using “evolving” scenes, ones that varied incrementally in clutter while maintaining their semantic continuity. Observers searched for a target building in rural, suburban, and urban city scenes created using the game SimCity. Stimuli were 30 screenshots obtained for each scene type as the city evolved over time. Reaction times and search guidance (measured by scan path ratio) were fastest/strongest for sparsely cluttered rural scenes, slower/weaker for more cluttered suburban scenes, and slowest/weakest for highly cluttered urban scenes. Subjective within-city clutter estimates also increased as each city matured and correlated highly with RT and search guidance. However, multiple regression modeling revealed that adding objective estimates failed to better predict search performance over the subjective estimates alone. This suggests that within-city clutter may not be explained exclusively by low-level feature congestion; conceptual congestion (e.g., the number of different types of buildings in a scene), part of the subjective clutter measure, may also be important in determining the effects of clutter on search.

Keywords: visual search, visual clutter, eye movements, feature congestion, edge detection

Citation: Neider, M. B., & Zelinsky, G. J. (2011). Cutting through the clutter: Searching for targets in evolving complex scenes. *Journal of Vision*, 11(14):7, 1–16, <http://www.journalofvision.org/content/11/14/7>, doi:10.1167/11.14.7.

## Introduction

Set size effects, the relationship between the number of objects in a display and the time needed to find a target, have long served as the gold standard for characterizing the efficiency of visual search (Wolfe, 1998b). Understanding this relationship is important, as it tells us how search is affected by load. From the many studies using this manipulation, we have learned that search efficiency often degrades roughly linearly as non-target objects are added to a display—the increased load arising from these added distractors makes it harder to find the search target. However, we also learned that this linearly increasing set size effect applies mainly to objects consisting of multiple features; when a target has a feature that is not shared with the distractors, its detection is often immediate regardless of the set size—a phenomenon commonly referred to as *pop out*. This observation led to early characterizations of search load effects in terms of dichotomies based on feature diversity and overlap, the most common being a relationship between parallel and serial search slopes to singleton and conjunction search tasks, respectively (Treisman & Gelade, 1980; see also Duncan & Humphreys, 1989, for a related distinction between homogeneous and heterogeneous displays). More recently, these dichotomy-based accounts have been subsumed under the broader framework of signal detection theory (Eckstein, 1998;

Palmer, Verghese, & Pavel, 2000; Verghese, 2001; but see Rosenholtz, 2001); as featurally diverse distractors are added to a display, noise is introduced that makes it more likely for non-target signals to be mistaken for the target, thereby degrading search efficiency.

In an effort to quantify search load effects in terms of distinct objects and features, studies using a set size manipulation have relied almost exclusively on relatively impoverished stimuli—arrays of simple visual patterns presented against a homogenous background (see Wolfe, 1998a, for a review). The advantages of using these stimuli are obvious; if the target is a red vertical bar and the distractors are red horizontal and green vertical bars, then the feature complexity and overlap between these objects can be precisely specified. Moreover, the use of simple and easily segregated stimuli makes quantifying search load trivial; to determine the set size, one needs only to count the number of objects.

Recent decades, however, have seen a growing push to use more ecologically valid stimuli in all corners of behavioral research, and visual search has been at the forefront of this ongoing, and indeed accelerating, trend. There are now many search studies that have used real-world targets displayed against a simple background (e.g., Biederman, Bickel, Teitelbaum, Klatsky, & Mezzanotte, 1988; Castelano, Pollatsek, & Cave, 2008; Newell, Brown, & Findlay, 2004; Schmidt & Zelinsky, 2009; Yang & Zelinsky, 2009), simple objects displayed against a

complex background (e.g., Brockmole & Henderson, 2006b; Wolfe, 1994b; Wolfe, Oliva, Horowitz, Butcher, & Bompas, 2002), real-world objects displayed against a complex background (e.g., Bravo & Farid, 2004; Neider, Boot, & Kramer, 2010; Neider & Zelinsky, 2006b), and of course realistic targets embedded in simple (e.g., Henderson, Weeks, & Hollingworth, 1999; Neider & Zelinsky, 2006a, 2010; Vö & Henderson, 2010; Zelinsky, 1999, 2001; Zelinsky, Rao, Hayhoe, & Ballard, 1997) or fully realistic scenes (e.g., Eckstein, Drescher, & Shimozaki, 2006; Foulsham & Underwood, 2007; Malcolm & Henderson, 2009; Oliva, Wolfe, & Arsenio, 2004; Zelinsky & Schmidt, 2009; see Henderson, 2003, 2007; Tatler, 2009, for reviews). Paralleling this barrage of behavioral studies has been an equally strong development in search theory, with many computational models of search now being able to accommodate realistic objects and scenes (e.g., Ehinger, Hidalgo-Sotelo, Torralba, & Oliva, 2009; Hwang, Higgins, & Pomplun, 2009; Itti & Koch, 2000; Kanan, Tong, Zhang, & Cottrell, 2009; Navalpakkam & Itti, 2005; Parkhurst, Law, & Niebur, 2002; Pomplun, 2006, 2007; Rao, Zelinsky, Hayhoe, & Ballard, 2002; Torralba, Oliva, Castelhana, & Henderson, 2006; Zelinsky, 2008).

This refocusing of the search literature from simple to realistic stimuli, however, has come with a price; the set size effect, our most accepted method of quantifying the effect of load on search, has become essentially meaningless. The problem lies in not knowing what counts as an “object” in the context of a real-world scene (Neider & Zelinsky, 2008, 2010; Rosenholtz, Li, Mansfield, & Jin, 2005; Wolfe, Vö, Evans, & Greene, 2011). Whereas the number of T and L letters in a standard search task is countable and yields a definitive answer, the number of distractors accompanying a coffee mug target in a typical kitchen scene is arguably unknowable. Although plates and other cups on the counter might constitute clear distractors, what about the toaster or the oven or every tile on the wall or pattern on the floor? Depending on how one chooses to define an object, the inclusion of these relatively non-object-like patterns in the count, or the parts of more accepted objects (each button on the blender or knob on the oven), might easily cause the estimated number of objects in a realistic scene to swell into the hundreds. Such arbitrariness corrupts an independent measure. Given that an “object” is the unit in a set size manipulation, as the meaningfulness of this unit breaks down, so does the hope of defining a set size.

As our hypothetical coffee mug search task exemplifies, all patterns in a realistic context are not likely to be treated equally by perception. It has long been known that some of these patterns are perceptually organized into “figures” or objects, with the rest delegated to the perceptual background (Wertheimer, 1923; for a more recent discussion and a review, see Craft, Schutz, Niebur, & von der Heydt, 2007; Driver, Davis, Russell, Turatto, & Freeman, 2001). This object/background division profoundly affects search. We explicitly demonstrated this in previous work

by creating camouflage backgrounds consisting of tiled pieces of the target object, essentially pitting the background against the object distractors in the display (Neider & Zelinsky, 2006b; see also Boot, Neider, & Kramer, 2009; Neider et al., 2010; Wolfe et al., 2002). Even though the distractors (other realistic objects) were clearly less featurally similar to the target than the background, they nevertheless attracted the vast majority of fixations during search. The fact that the search process segregated objects from background, even under conditions of camouflage, led us to speculate that a similar segregation may also characterize search through fully realistic scenes (Neider & Zelinsky, 2010). If there is any hope of defining a set size effect for such scenes, it is imperative that this process be understood, as it would be the objects segregated from the background that would impose the load and impact search efficiency.

Even if this segregation process can be deciphered and the objects in a scene counted, there is no guarantee that these objects would comprise a stable set—what may be considered an object in one search context or scene may not be considered an object in another. In Neider and Zelinsky (2008), we showed that the definition of an object, and therefore the scene’s *relevant set size*, might change with the number of countable distractors in a scene. Using scenes varying in their number of tree distractors, we found that at low set sizes the objects indeed seemed to be the trees, resulting in standard set size effects. However, at higher set sizes, which allowed for the trees to become grouped into clumps, observers redefined the scene’s objects to be the open field regions emerging between the clumps of trees. As distractors were added to a scene, the observer changed the definition of what counted as an object from “tree” to “field.” When we plotted search time as a function of the actual number of tree distractors, this object redefinition produced a negative set size effect, a pattern that is rarely found in studies using simple stimuli but one that might actually be common in the case of realistic scenes. All of this casts doubt on whether common conceptions of a set size effect can be meaningfully extended into the real world. If an object is not a static entity, but rather a perceptual construct that can change at the discretion of the observer, how then is it possible to count the number of objects appearing in a scene? Determining a set size for realistic scenes might therefore not be just a difficult problem, it might be an ill-defined problem.

If standard conceptions of set size cannot be applied to realistic search tasks, how then is it possible to quantify effects of load on visual search in these tasks? One intriguing suggestion by Rosenholtz et al. (Rosenholtz et al., 2005; Rosenholtz, Li, & Nakano, 2007) is to use visual clutter as a surrogate measure of set size in realistic scenes. Even if the objects in a scene are uncountable, these objects are composed of features, and the “congestion” among these features can be quantified. The expectation is that feature congestion should increase with the number

of objects, regardless of how they may be defined, and therefore predict search efficiency. Rather than quantifying search efficiency in terms of a  $RT \times$  set size function, changes in search efficiency with increasing load might be quantified in terms of a  $RT \times$  clutter function in the context of scenes.

How should feature congestion be quantified? Recent work by Henderson, Chanceaux, and Smith (2009) evaluated several image-based measures of visual clutter in the context of a search task, including edge density, sub-band entropy, and feature congestion. Feature congestion was quantified using the Feature Congestion model (Rosenholtz et al., 2005, 2007), which computes local variability in color, orientation, and luminance contrast over the whole of an image to derive a measure of visual clutter for that image (also see Bravo & Farid, 2008, for a related method of quantifying feature clutter). Although all of these clutter measures were found to correlate reasonably well with search RTs, perhaps surprising was the finding that one of the simpler measures, edge density, predicted search performance about as well as the rest. As the number of edges in a scene increased, so did search RTs, perhaps due to a relationship between the proximity of neighboring edges and one's ability to segment a scene into objects (Bravo & Farid, 2004, 2006). In addition to their computational simplicity and incontrovertible inclusion in the pantheon of early visual features, edges have also been shown to correlate highly with gaze fixations in both simple displays (Mackworth & Morandi, 1967) and real-world scenes (Baddeley & Tatler, 2006), a relationship of particular relevance for the present study.

Despite this flurry of recent research relating clutter to search performance, there are two aspects of this relationship that deserve further consideration. First, while it has been shown that search RTs tend to increase with the visual clutter of real-world scenes (e.g., Henderson et al., 2009; Rosenholtz et al., 2005), this has mainly been demonstrated in the context of unrelated search displays. In a typical study, observers are shown a scene and asked to search for a target, but a qualitatively different scene is used on every trial (e.g., a street scene followed by a living room scene, etc.). These scenes are then quantified in terms of clutter and correlated with search performance. Such a lack of continuity across scene stimuli introduces several sources of variability that are typically absent or minimized in more traditional set size manipulations. For example, random scenes not only have different levels of clutter, they will also have different underlying visual statistics (Greene & Oliva, 2009; Oliva & Torralba, 2006). Some scenes are dominated by vertical features and greenish hues (e.g., forest scenes), while others are dominated by horizontal features and bluish hues (e.g., ocean scenes). These scene-specific features, to the extent that they differ in their similarity relationships to the target, would affect search performance and ultimately lower any correlation between search and clutter. Moreover, random scenes

will have different semantic labels (e.g., “forest” or “ocean”), and this too would be expected to affect search via the introduction of contextual constraints (Torralba et al., 2006). Neither of these sources of variability are meaningfully present in a typical set size manipulation. Increasing the number of rotated L distractors in a T search task imposes a greater load, but the underlying “scene” does not change. What is needed is a search task in which clutter varies from trial to trial while other factors, such as feature heterogeneity and scene semantics, are held relatively constant.

Other studies have used fairly homogenous stimulus sets, such as maps (Rosenholtz et al., 2007) or the contents of handbags (Bravo & Farid, 2008), that have largely avoided this problem, but once again the claim that increasing clutter is analogous to increasing set size was not explicitly tested. Central to the concept of a set size manipulation is the incremental addition of distractors to a scene. Although map and handbag stimuli might all have the same “map” and “handbag” semantic labels, individual elements were not systematically added to the map displays nor were items incrementally inserted into the handbag scenes. These semantically homogenous clutter stimuli therefore more closely approximate standard search experiments in which different combinations of letter stimuli are used as targets and distractors, such as a T in Ls or an O in Qs, but there is nothing akin to a set size manipulation. In the absence of an incremental manipulation of clutter, the question of whether clutter can serve as a surrogate measure of search set size effects remains largely unanswered.

To address this need, in the present study, we used commercially available gaming software to create scenes of cities that evolved over time. Each of these search scenes started with the same “base scene” (time 1), which depicted a largely barren field with a few scattered roads. From this common origin, we then developed three types of cities: a rural city, a suburban city, and an urban city. We did this by investing different levels of resources within the context of the game, resulting in highly cluttered urban scenes, less cluttered suburban scenes, and sparsely cluttered rural scenes. We predicted that the additional structures needed to transform a rural city into a suburban city, and a suburban city into an urban city, would add clutter and, consequently, decrease search performance in between-city comparisons. By the same logic, because a city at time 1 should be less cluttered than the same city at time 10, which should be less cluttered than at time 20, these within-city comparisons might also reveal deteriorating search performance with increasing subjective and objective clutter. Furthermore, the expectation that clutter will accumulate more quickly during the evolution of an urban city compared to a suburban or rural city leads us to predict a scene type  $\times$  within-city clutter interaction, with a fully matured urban city showing the highest level of clutter and the worst search performance. Importantly, all



of these predictions are made in the context of semantically related scenes that all evolved from the same base scene, and in the context of an incremental manipulation of clutter, one that reasonably approximates a standard set size manipulation.

Second, previous studies relating clutter to search have relied almost exclusively on purely objective techniques to quantify visual clutter, such as counting edges or estimating feature congestion (but see Beck, Lohrenz, & Trafton, 2010; Lohrenz, Trafton, Beck, & Gendron, 2009; Rosenholtz et al., 2005; van den Berg, Cornilissen, & Roerdink, 2009). Although objective clutter estimates are important and are producing interesting results, perhaps equally important is the collection of *subjective* clutter estimates. Clutter is, after all, a percept, and percepts are largely subjective in nature—what one person perceives as cluttered another person might perceive as sparse. Subjective clutter estimates might also be influenced by a host of top-down factors, as opposed to bottom-up factors derived solely from the search image. Such factors might affect search behavior by altering the representation of the target (Chen & Zelinsky, 2006; Yang & Zelinsky, 2009; Zelinsky, 2008, Experiment 3) or by introducing semantic associations or contextual scene constraints that restrict the search space (e.g., Biederman, Glass, & Stacy, 1973; Brockmole & Henderson, 2006a, 2006b; Eckstein et al., 2006; Henderson et al., 1999; Neider & Zelinsky, 2006a; Torralba et al., 2006; Zelinsky & Schmidt, 2009; see Oliva & Torralba, 2007, for a review). The potentially considerable variability in search behavior arising from these top-down factors would not be captured by purely bottom-up, objective clutter estimates. This may explain why these purely objective estimates correlate only modestly with search performance (e.g.,  $R^2$  on the order of  $\sim 0.3$  for several measures in Henderson et al., 2009). In the absence of subjective clutter estimates, it is therefore impossible to know whether these modest correlations are due to objective clutter estimates failing to characterize search behavior specifically or the perception of clutter more generally.

We address this problem by collecting subjective clutter ratings for rural, suburban, and urban scene types at each step in their development over time. By doing this, we create a sort of psychological ground truth—a measure of how cluttered a scene is perceived to be. These subjective estimates also enable an intermediate step in the evaluation of existing objective clutter techniques; objective estimates can be correlated with the subjective estimates as well as the ultimate search performance. It may be that objective clutter estimates correlate highly with subjective clutter estimates for the same scenes but that this correlation drops off when extended to actual search behavior. Finding this pattern would suggest that objective estimates are indeed valuable in capturing perceived clutter but that they are limited as a description of search due to their failure to account for top-down factors affecting

search performance. Alternatively, we might find that these objective estimates correlate only modestly with both search performance *and* subjective assessments of clutter. This pattern would indicate a more profound limitation of the objective clutter technique. In the context of our evolving quasi-realistic scenes, we might also find objective clutter correlating well with subjective estimates and search behavior. This pattern would suggest that the previously reported modest correlations were likely due to variability introduced by the use of random unrelated scenes. For the sake of completeness, is it also possible that objective clutter might correlate poorly with subjective clutter but highly with search, although we consider this possibility to be unlikely.

## Methods

### Stimuli and design

Scenes varying in clutter were created using the video game SimCity 4 (EA Games, 2003). SimCity 4 is a civic simulator that allows players to create unique virtual cities that grow over time according to the game's simulation engine. Three scene types were created: rural, suburban, and urban cities (Figure 1). In accordance with the realistic civic planning rules incorporated into the game, the rural city was constructed using low-density residential, commercial, and industrial zoning, and the suburban and urban cities were constructed using medium-density and high-density zoning, respectively. These different zoning restrictions constrained the types of structures that the game could use to build the cities, ultimately producing cities that varied in their degree of clutter. Importantly, the construction of each city from the same starting base landscape imposed a degree of visual and contextual self-similarity on these scenes, meaning that comparisons between these city scenes would be more likely to reflect pure differences in clutter—at least compared to random scenes. Our expectation was that the rural city would be the least cluttered, the suburban city would have an intermediate level of clutter, and the urban city would be the most cluttered. Moreover, the fact that each city had a common origin means that these expected clutter differences should emerge only after the cities had an opportunity to evolve. Individual scenes were selected so as to capture this evolution. We captured images of each city at 30 fixed points in time during its growth, resulting in a total of 90 rural, suburban, and urban scenes that incrementally increased in clutter. Images captured at early time points were generally sparse, as the cities would not have undergone much development, whereas images captured at later time points appeared denser, reflecting the maturation of the city.

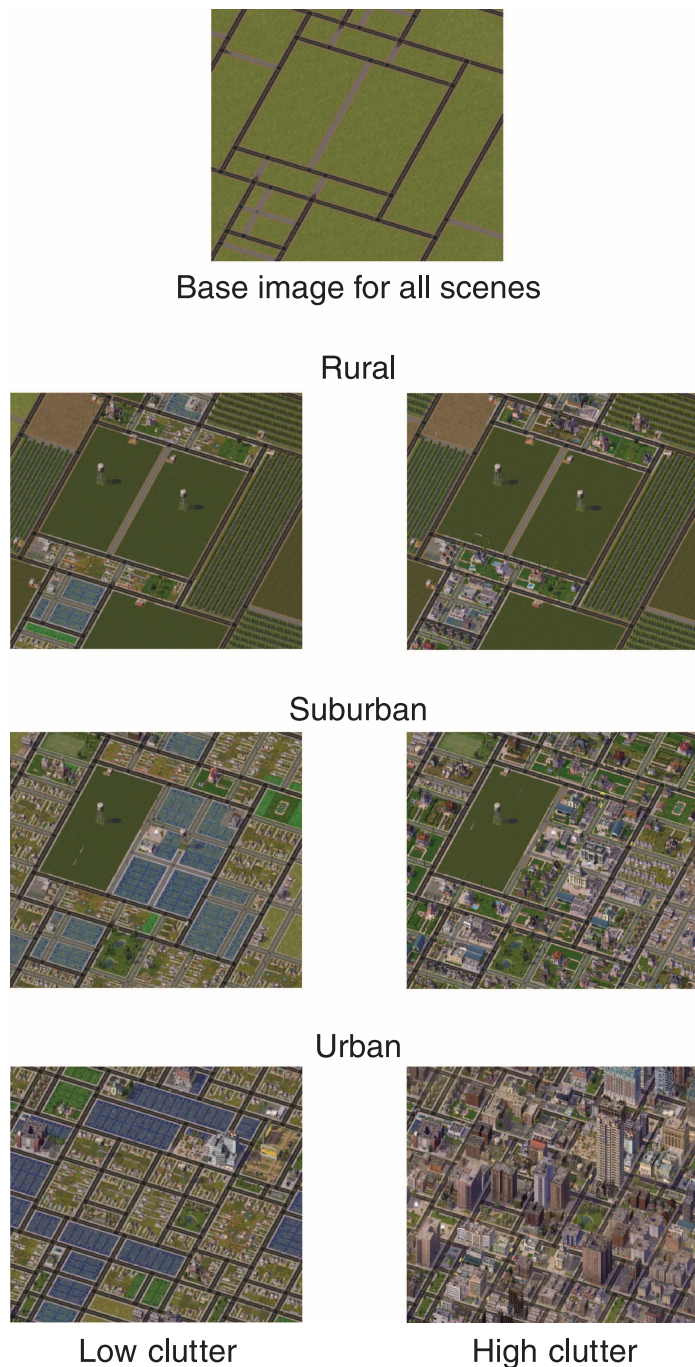


Figure 1. Each of the three scene types started from the same base image but then matured during game play into more typical depictions of rural, suburban, and urban cities. Low-clutter scenes were captured early during game play; high-clutter scenes were captured later during game play.

### Subjective clutter rating procedure

We used a multistage rating procedure to obtain a subjective measure of visual clutter for these scenes and to validate the expected clutter changes resulting from normal city evolution during game play. First, the 30 images

from each scene type were printed in color, shuffled so that the images were in a random order rather than the order in which they were produced by the game, and placed in separate folders. Twenty-four Stony Brook University undergraduate students (none of whom participated in the search experiment) then rank ordered these images, lowest to highest, for visual clutter. This was done for all three scene types, rural, suburban, and urban, in a blocked and counterbalanced order. From these individual orderings, we then calculated a median ordering of images for each scene type, reflecting our raters' average perception of within-city clutter relationships.

Following this initial rating stage, twelve new raters assigned a *clutter transition score* to each consecutive image pair within each scene's 30 rank-ordered images (from the first stage). This was done so as to capture pairwise magnitude differences in subjective clutter in our three median-ordered lists, thereby enabling us to estimate the degree that clutter is perceived to change from scene to scene throughout a city's evolution. To generate these estimates, each rater assigned a score between 0 and 10 indicating the perceived magnitude of the clutter change between images  $n$  and  $n + 1$ . For example, a given rater might have assigned a score of 6 to the transition between images 3 and 4 in the ordered set of urban scenes but assigned a score of 2 to the transition between images 20 and 21. This process was continued until each consecutive image pair was associated with a transition score. A subjective clutter estimate for each of the 30 images in each of our three city scene types was then derived by adding the median clutter transition score to the median clutter score from the previous image pairing; the first image in each set, which was identical across scene type, was assigned a score of 0.

### Search task procedure and apparatus

Twelve experimentally naive Stony Brook University undergraduates, all of whom had normal or corrected-to-normal vision (by self-report), participated as part of a course requirement. Search stimuli were constructed from the above-described set of 90 scenes. Appearing in each of these scenes was a town hall building (Figure 2), which was the designated target of the search task. To force the appearance of this building in each city, we used a provision of the game allowing the player to request the construction of a particular structure. The game therefore inserted the target into each scene; targets were not digitally inserted after the fact by the experimenter. This is significant, as it removes the concern that artifacts related to image manipulation might have affected gaze behavior during search. The target building was  $\sim 2.86^\circ$  along its largest dimension and was located equally often in each of the four image quadrants for each scene type.

Two search scenes were created from each captured image (with different target placements), yielding 180 unique

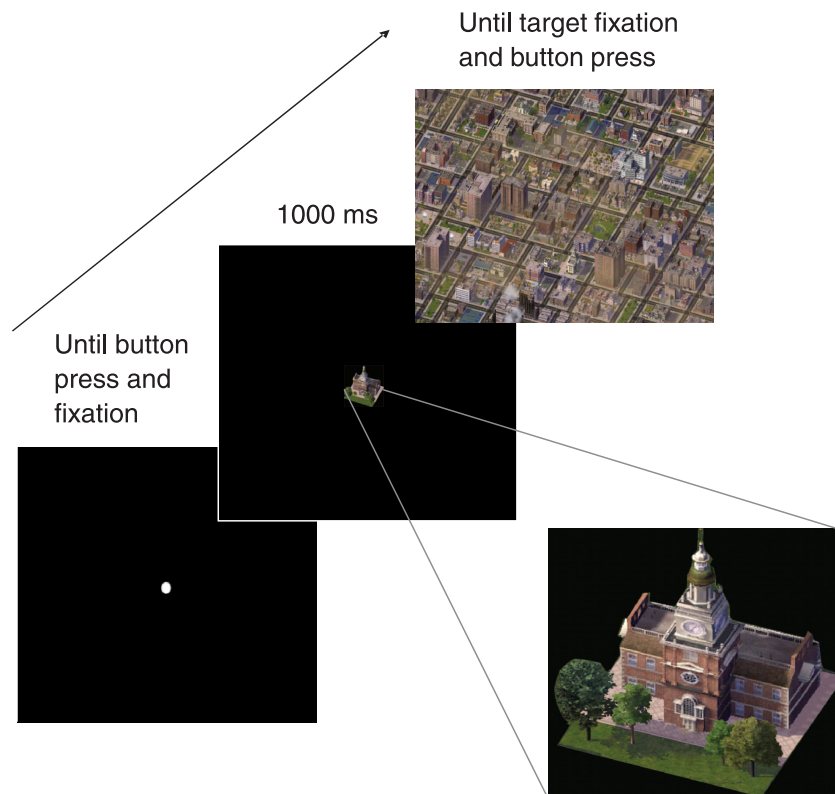


Figure 2. Procedure used in the search experiment.

scenes (60 per city type). Note however that because all three cities used a common starting scene, this design would produce 6 largely identical images at game time 1; rural, suburban, and urban cities would not yet have differentiated, meaning only the target placement would differ. To avoid this redundancy, we presented scene 1 only twice to observers, thereby maintaining an equal number of presentations of each unique scene. As a result of this exclusion, we were left with 176 unique scenes for use as search stimuli. Each search scene subtended  $27^\circ \times 20^\circ$  of visual angle and was presented in color on a 19" CRT monitor. Eye movements were recorded throughout using an SR Research EyeLink II eye tracker sampling at 500 Hz (with chin rest). All eye movement measures were quantified using the tracker's default algorithms and settings. Button presses were collected from the observer's preferred hand using the two triggers of a GamePad controller.

The events comprising a typical trial are illustrated in Figure 2. Each trial began with the presentation of a fixation dot at the center of the screen. Upon its fixation, and a manual button press (for drift correction), a picture of the target building was presented centrally for 1 s, followed by the search scene. The target cue was the same on every trial, and every search scene contained the target exactly as it had appeared in the cue. The observer's task was to find and fixate the target and to press a button while holding gaze on the object. In the case of a false alarm (the observer pressing the button while their gaze was not on the target), a tone sounded indicating to the

observer that they had not accurately located the target and that they should continue searching. There were 6 practice trials, followed by 176 experimental trials.

## Results

### Comparing subjective and objective clutter

Pronounced differences in subjective clutter scores were found across the three scene types,  $F(2, 58) = 105.82$ ,  $p < 0.001$ . Bonferroni corrected post hoc comparisons revealed that urban scenes (mean = 14.49) were assigned higher clutter scores than both suburban (mean = 9.29,  $p < 0.001$ ) and rural (mean = 7.63,  $p < 0.001$ ) scenes, with suburban scenes also being scored as more highly cluttered than rural scenes ( $p < 0.001$ ). We also analyzed how clutter within each scene type changed with the city's evolution (e.g., the clutter of urban scene  $n$  compared to urban scene  $n + 1$ ). As expected, clutter scores were highly correlated with city maturation in all three scene types ( $R^2 = 0.97, 0.96$ , and  $0.98$  for rural, suburban, and urban scenes, respectively; all  $p < 0.001$ ); as in the case of a set size manipulation, as each city grew, observers rated these scenes as becoming more cluttered.

We explored two objective measures of visual clutter. First, we calculated the density of edges in each of the city



scenes. This was done using the Canny edge detection method (Canny, 1986) implemented in Matlab (v. 7.8.0). The Canny method defines strong and weak edges as local maxima in the intensity gradient of a grayscale image and includes weak edges only if they are connected to a strong edge.<sup>1</sup> Second, we obtained clutter estimates using the feature congestion model (Rosenholtz et al., 2007). Our implementation of this model used Matlab code made publicly available by the first author. Rosenholtz et al. (2007) should be consulted for further details regarding the feature congestion model. Although we analyzed clutter using each of these objective methods, comparison of the results revealed a high degree of redundancy in the obtained patterns. Throughout the remainder of the paper, we therefore report results only from the edge density estimates of objective clutter; parallel analyses using estimates from the feature congestion model can be found in [Supplementary materials](#).

As in the case of subjective clutter, edge clutter differed significantly between the three scene types,  $F(2, 58) = 551.59$ ,  $p < 0.001$ . Urban scenes (mean = 33,264) had higher edge densities than both suburban (mean = 28,933,  $p < 0.001$ ) and rural scenes (mean = 18,687,  $p < 0.001$ ), with suburban scenes having more edges than rural scenes ( $p < 0.001$ ). Edge density also correlated with city maturation ( $R^2 = 0.39, 0.46, \text{ and } 0.16$  for rural, suburban, and urban scenes, respectively; all  $p < 0.05$ ); as each city grew, so too did edge clutter. Together, these analyses serve to validate the use of these scenes in studies of visual clutter. Regardless of whether clutter is measured subjectively by observer ratings or objectively in terms of edges, these three cities differed in terms of clutter, with clutter increasing as each city matured.

To determine whether the objective count of edges in a scene captured the subjective perception of scene clutter, we correlated the subjective clutter scores with edge

density in the 30 rural, 30 suburban, and 30 urban images used in the subjective rating task. Previous work has shown edge density to be a reasonably good predictor of visual clutter effects on search performance (Henderson et al., 2009; Rosenholtz et al., 2007), with the assumption being that edge density would also correlate with the subjective perception of clutter. Our analysis largely confirmed this relationship; edge density correlated highly with subjectively perceived visual clutter for all three of our scene types,  $R^2 = 0.56, 0.62, \text{ and } 0.26$  in the rural, suburban, and urban cities, respectively (Table 1). This suggests that the objective method of counting edges in a scene may be a reasonable way of characterizing the subjective perception of scene clutter, at least for the evolving city scenes used in this study. However, the fact that these correlations are imperfect means that other factors not captured by edge density (or feature congestion; see [Supplementary materials](#)) also contribute to subjective clutter percepts. In the following sections, we relate both objective edge density and subjective clutter estimates to search behavior and attempt to determine which is the better predictor of performance in our search task.

## Manual error rates

Trials in the search task terminated only after an observer fixated the target while pressing a button indicating that they had located the target, so a true error could not be made in this task. However, observers could mistake non-targets for the target, as indicated by a button press response when some item other than the target was fixated. These false alarm rates averaged 1.9%, 1.1%, and 1.5% in rural, suburban, and city scenes, respectively, and did not differ significantly across scene type,  $F(2, 22) = 1.65$ ,

	Subjective ratings	Edge density	RT/log(RT)	Scan path ratio
<b>Rural</b>				
Subjective ratings	1.0	0.56***	0.52***/0.56***	0.30**
Edge density		1.0	0.30**/0.44***	0.16*
RT/log(RT)			1.0/1.0	0.60***/0.57***
Scan path ratio				1.0
<b>Suburban</b>				
Subjective ratings	1.0	0.62***	0.53***/0.71***	0.21*
Edge density		1.0	0.31**/0.61***	0.12
RT/log(RT)			1.0/1.0	0.65***/0.44***
Scan path ratio				1.0
<b>Urban</b>				
Subjective ratings	1.0	0.26*	0.53***/0.74***	0.49***
Edge density		1.0	0.10/0.33*	0.12
RT/log(RT)			1.0/1.0	0.83***/0.74***
Scan path ratio				1.0

Table 1. Correlations ( $R^2$ ) between reaction time (RT), log(RT), scan path ratio, and clutter estimates by scene type. Notes: \* $p < 0.05$ ; \*\* $p < 0.005$ ; \*\*\* $p < 0.001$ .



$p = 0.22$ . Trials in which a false alarm occurred were excluded from all subsequent analyses.

## Manual reaction times

Reaction times (RTs) generally increase with set size in most complex search tasks—does a similar relationship exist between search RT and visual clutter? One way to answer this question is to compare RTs between scene types that differ in both subjective and objective clutter. Analysis of variance revealed clear between-scene effects of clutter on RT,  $F(2, 58) = 26.55$ ,  $p < 0.001$  (Table 2). Post hoc comparisons confirmed that targets in urban scenes took longer to find than in suburban ( $p < 0.005$ ) or rural ( $p < 0.001$ ) scenes, with search in the suburban scenes also requiring more time compared to the less cluttered rural scenes ( $p < 0.001$ ). This finding replicates previous work showing longer search times for more cluttered visual scenes (Henderson et al., 2009; Rosenholtz et al., 2007).

To further specify this relationship, we plot, in Figure 3a, RT as a function of the subjectively rated visual clutter scores assigned to each of the images from the three scene types. Reaction times were found to increase with within-city clutter; regardless of scene type, the increased subjective clutter accompanying city maturation could account for about 53% of the variance in manual search times. Analysis of the slopes of the best-fit regression lines for each scene type revealed that search times increased with within-city clutter faster in the urban scenes than in the suburban,  $t(11) = 4.34$ ,  $p < 0.005$ , or rural scenes,  $t(11) = 7.13$ ,  $p < 0.001$ , and that this increase was also faster for suburban scenes compared to rural scenes,  $t(11) = 6.59$ ,  $p < 0.001$ . These slope differences reflect a scene type  $\times$  within-city clutter interaction in search times; targets took longer to find in scenes that became more subjectively cluttered as they matured.<sup>2</sup> Correlations between subjective clutter and multiple measures of search performance appear in Table 1, including RT and  $\log(\text{RT})$ . This latter measure was included to assess the potential for a non-linear relationship between RT and our clutter estimates.<sup>3</sup>

Figure 3b shows a similar relationship between RT and edge density, one of our objective estimates of visual clutter (see also Supplementary materials). We again found that search RTs were positively correlated with visual clutter in the rural and suburban scenes, with edge density accounting for roughly 30% of the variability in

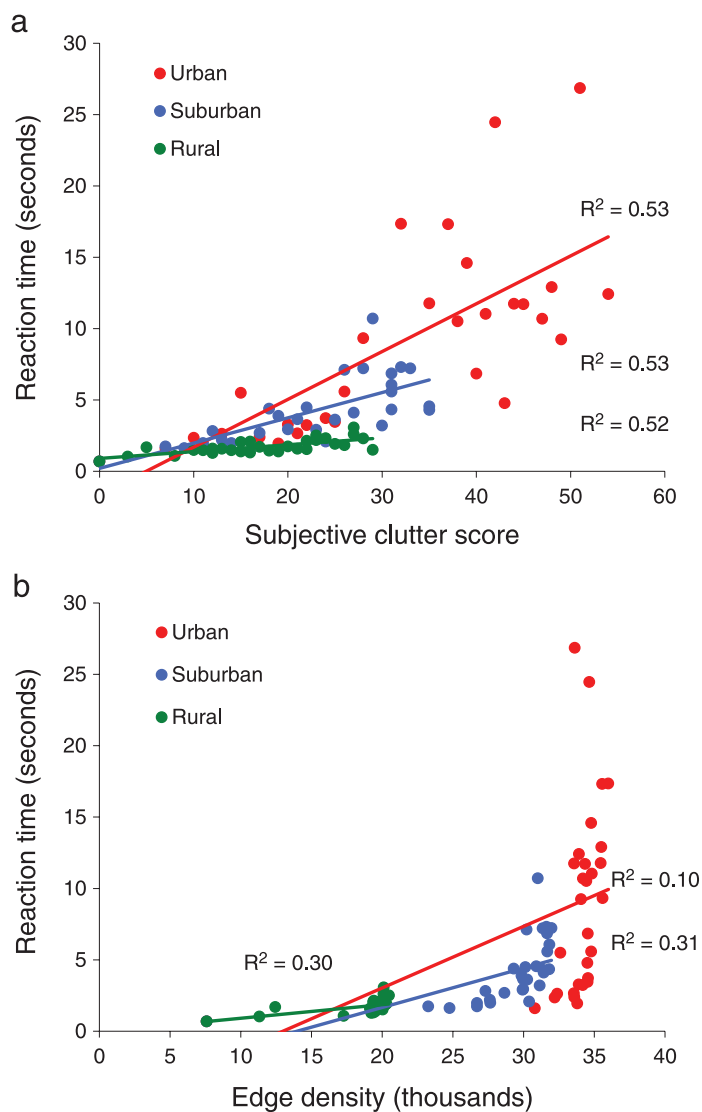


Figure 3. Reaction times and best-fit regression lines for urban, suburban, and rural scenes as a function of (a) subjective clutter score and (b) edge density. See Supplementary materials for a corresponding plot showing the relationship between reaction time and feature congestion.

search times for each task (Table 1). Comparing the slopes of the best-fit regression lines again confirmed that RTs increased with edge clutter faster in the suburban scenes compared to the rural scenes,  $t(11) = 6.76$ ,  $p < 0.001$ . However, the apparently steeper slope found for urban

	Reaction time (ms)	Scan path ratio	Final saccadic amplitude (degrees)	Target verification time (ms)
Rural	1730 (93)	4.09 (0.21)	2.44 (0.19)	550 (18)
Suburban	4130 (415)	8.92 (0.47)	2.01 (0.18)	670 (25)
Urban	8752 (1221)	16.29 (1.52)	1.55 (0.12)	729 (28)

Table 2. Mean search performance measures by scene type. Notes: Values in parentheses indicate one standard error of the means.

scenes should be treated with caution, as the correlation between RTs and edge density for this scene type was weaker ( $R^2 = 0.1$ ) and only trended toward significance ( $p = 0.08$ ). Although search times increased with edge content in the case of rural and suburban scenes, in the case of urban scenes, adding edges did not result in reliably longer RTs.

From the above analyses, we know that search difficulty generally increased with both subjective and objective visual clutter, but do these two measures capture different aspects of clutter's influence on search performance? To address this question, we conducted a multiple regression analysis for each scene type, with subjective clutter ratings and edge density being predictor variables and RT being the dependent variable. When variability due to edge density was partialled out of the  $RT \times$  subjective clutter correlations, we still found highly significant correlations for all three scene types ( $R_{\text{partial}} = 0.555, 0.559,$  and  $0.689$  for the rural, suburban, and urban scenes, respectively; all  $p < 0.005$ ). However, when we removed variability associated with subjective clutter from the objective estimates, these correlations disappeared ( $R_{\text{partial}} = 0.015, -0.041,$  and  $-0.076$  for the rural, suburban, and urban scenes, respectively; all  $p > 0.9$ ). The same regression analyses conducted for  $\log(\text{RT})$  produced a similar pattern of results. This asymmetrical relationship suggests that nearly all of the variability in search times attributable to edge clutter can be accounted for by our subjective clutter estimates but that the converse is not true; objective clutter estimates based on edge density fail to capture aspects of the relationship between clutter and manual search efficiency that is captured by our subjective estimates (see [Supplementary materials](#) for an identical analysis using feature congestion rather than edge density, with identical conclusions).

## Eye movement guidance to the target

Reaction times in a search task can be meaningfully decomposed into a guidance component, an observer's efficiency in moving their eyes to a target, and a decision component that captures the time needed to reject fixated distractors or verify the presence of a target (e.g., Castelano et al., 2008; Malcolm & Henderson, 2009; Yang & Zelinsky, 2009). To characterize the relationship between visual clutter and search guidance, we computed a scan path ratio for each trial (Castelano & Henderson, 2007). This ratio is calculated by dividing the distance traversed by the eyes during a trial (summed Euclidean distance between fixations 1 ...  $n$ , where  $n$  is the first fixation on the target) by the most efficient possible route to the target, as quantified by the distance from the center of the screen to the center of the target. A scan path ratio of 1 would, therefore, indicate a direct path to the target and maximal guidance, with values greater than 1 indicating increasingly inefficient paths and weaker guidance.

The results of these analyses are reported in [Table 2](#). As in the case of manual RTs, scan path ratios varied across scene type,  $F(2, 22) = 55.15, p < 0.001$ . In the less cluttered rural scenes, observers moved their eyes toward the target far more efficiently (4.09) than in the suburban (8.92;  $p < 0.001$ ) or urban (16.29;  $p < 0.001$ ) scenes. Scan paths were also more efficient in the suburban scenes compared to the urban scenes,  $p < 0.001$ . Overall, as between-scene visual clutter increased, observers moved their eyes less directly to the target.

To examine the effect of within-city clutter on search guidance, we correlated each observer's scan path ratio on

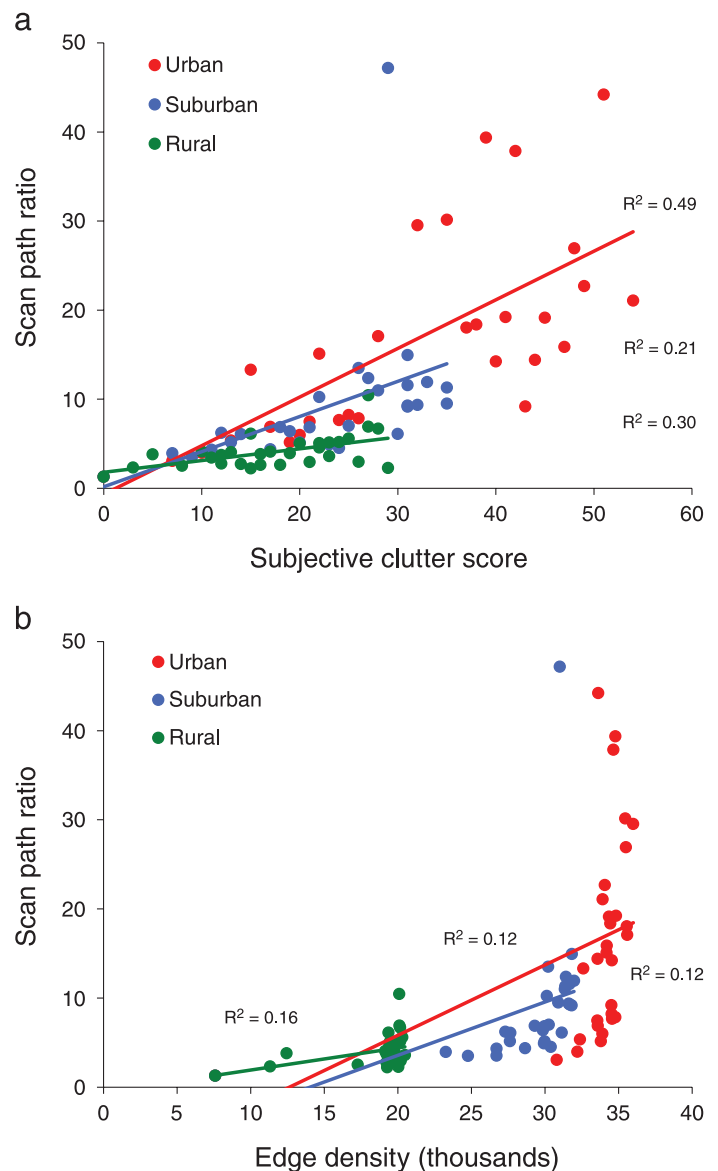


Figure 4. Scan path ratios and best-fit regression lines for urban, suburban, and rural scenes as a function of (a) subjective clutter score and (b) edge density. See [Supplementary materials](#) for a corresponding plot showing the relationship between scan path ratio and feature congestion.

an image-by-image basis with the subjective (Figure 4a) and objective (Figure 4b) clutter estimates associated with that particular image (also see Table 1). In general, scan path ratios correlated well with subjective clutter scores ( $R^2 = 0.30, 0.21,$  and  $0.49$  in the rural, suburban, and urban scenes, respectively; all  $p < 0.05$ ), with the slopes of the best-fit regression lines suggesting that search efficiency decreased with accumulating subjective clutter fastest in the urban scenes and slowest in the rural scenes (all  $p < 0.05$ ). Correlations with within-scene clutter as measured by edge density were less impressive (see also Supplementary materials). While scan path ratios correlated with edge density in rural scenes ( $R^2 = 0.16; p < 0.05$ ), correlations with suburban ( $R^2 = 0.12; p = 0.06$ ) and urban scenes ( $R^2 = 0.12; p = 0.06$ ) only trended toward significance. Multiple regression analyses again showed that removing variability due to edge clutter from the scan path  $\times$  subjective clutter correlations still produced reliable correlations for the rural and urban scene types ( $R_{\text{partial}} = 0.42$  and  $0.65$ , respectively; both  $p < 0.05$ ) and a correlation for the suburban scenes that trended toward significance ( $R_{\text{partial}} = 0.33, p = 0.08$ ). However, the converse relationship again failed to hold; removing the factor of subjective clutter from the scan path  $\times$  objective clutter correlations resulted in the loss of all three correlations ( $R_{\text{partial}} = -0.03, -0.04,$  and  $-0.01$  for the rural, suburban, and urban scenes, respectively; all  $p > 0.85$ ). Visual clutter as measured by edge density did not account for additional variance in search guidance over and above that accounted for by the subjective ratings of clutter.

### Amplitude of final saccade to the target

To examine how clutter affects local detectability, we analyzed the amplitudes of the saccades that brought gaze to the target object (Table 2). Final saccade amplitude has been used as a measure of an observer's ability to detect a target pattern in their visual periphery (e.g., Engel, 1977; Henderson et al., 1999; Krendel & Wodinsky, 1960), with longer saccades indicating a greater distance over which peripherally viewed targets can be detected. We found that saccades to targets differed in amplitude as a function of between-city clutter,  $F(2, 22) = 16.71, p < 0.001$ ; observers made longer final saccades in rural ( $2.44^\circ$ ) and suburban ( $2.01^\circ$ ) scenes than in urban scenes ( $1.55^\circ$ ; both  $p < 0.005$ ) and in rural scenes compared to suburban scenes ( $p < 0.05$ ). However, final saccade amplitude was a relatively weak predictor of within-city clutter, correlating significantly with subjective clutter for urban scenes ( $R^2 = 0.14; p < 0.05$ ) but not for rural ( $R^2 = 0.02; p = 0.47$ ) or suburban ( $R^2 = 0.0004; p = 0.89$ ) scenes. Final saccadic amplitude failed to correlate reliably with edge density for any scene type (all  $p > 0.49$ ; see also Supplementary materials).

### Target verification time

Scan path ratio and final saccadic amplitude characterized how visual clutter affected eye movement guidance during search, but it is also possible that visual clutter affected the search decision following fixation of the target. If this was the case, then we might expect observers to have looked longer at the target under conditions of increasing clutter before their button press response indicating target detection. These target verification times, defined as the time between the initial fixation on the target and the search judgment, are shown in Table 2. As in the case of the guidance measures, target verification times varied across scene type,  $F(2, 22) = 65.37, p < 0.001$ . Observers needed less time to make their detection decisions for the relatively uncluttered rural scenes (550 ms) compared to the more cluttered suburban (670 ms;  $p < 0.001$ ) and urban (729 ms;  $p < 0.001$ ) scenes. Similarly, observers were slower to verify that a fixated object was the target in urban scenes than in the less cluttered suburban scenes,  $p < 0.005$ .

This relationship between between-city clutter and verification time extended also to within-city clutter. Target verification times correlated with within-city clutter as estimated by both subjective ratings ( $R^2 = 0.22, 0.18,$  and  $0.39$  for rural, suburban, and urban scenes, respectively; all  $p < 0.05$ ) and edge density ( $R^2 = 0.29, 0.32,$  and  $0.34$  for rural, suburban, and urban scenes, respectively; all  $p < 0.01$ ). However, and unlike the dependent measures reflecting search guidance, multiple regression analyses revealed no clear superiority of the subjective clutter estimates over the objective clutter estimates as a predictor of verification difficulty. The increase in visual clutter that occurs as a city matures makes it more difficult for observers to verify the detection of a search target, and both types of estimates were important in capturing this effect of visual clutter.

## General discussion

We set out to accomplish two goals in this study. First, we wanted to develop a test bed of scene stimuli that vary incrementally in clutter, thereby minimizing the large differences in image statistics that may exist when semantically unrelated scenes are used as stimuli. Given that a set size manipulation describes a similarly incremental change as distractors are added to a display, this enables us to better determine whether clutter can be used as a surrogate measure of set size in realistic scenes. Second, we wanted to compare both objective (edge density and feature congestion) and subjective (rater scores) estimates of clutter to search performance. Finding that search behavior correlates with both types of estimates, and

that each correlates with the other, would mean that search load effects can be measured using current objective methods. However, finding that only the subjective measure of clutter correlates well with search would suggest that semantic factors may need to be added to existing models of clutter before their estimates can be used as a surrogate measure of search set size.

With regard to the relationship between objectively defined clutter and search, previous work found that search performance tends to deteriorate as visual clutter increases; targets take longer to find in more cluttered scenes (e.g., Henderson et al., 2009; Rosenholtz et al., 2005). The use of unrelated scenes as stimuli, however, raised the possibility that the very modest correlations reported between objective clutter and search might be due to visual factors related to the semantics of the scenes masking a stronger relationship. Our method of evolving scenes ensured that all of our stimuli were highly related, both visually and semantically, certainly far more so than random real-world scenes. We found that manual search times increased with objective clutter between rural, suburban, and urban city scenes, a pattern consistent with previous studies that have used related scenes as search stimuli (Bravo & Farid, 2008; Rosenholtz et al., 2007). Moreover, the magnitude of these clutter effects were comparable to previous reports (Henderson et al., 2009), suggesting that the variability between random real-world scenes does not dampen the relationship between search and objective clutter; even when this variability is minimized, we find that this relationship does not improve. The incremental nature of the city maturation process also enabled us to test claims regarding the use of clutter as a surrogate measure of search set size in complex scenes. Although we found within-scene correlations between objective clutter and RT for rural and suburban cities, these relationships were modest at best ( $R^2 \sim 0.30$ ), and no reliable correlation was found for the most densely cluttered urban city. Edge density and feature congestion are therefore adequate to capture relatively large differences in clutter, as exemplified by the reliable and robust between-city effects reported in this paper, but more powerful objective methods are needed to capture the incremental changes in clutter that most closely approximate a set size manipulation.

In contrast to the modest correlations found between objective clutter and manual search performance, correlations between subjective clutter estimates and search were typically much stronger and more robust. Regardless of the scene type, subjective clutter could account for roughly half ( $R^2 = \sim 0.52$ ) of the variability in manual search RTs, about 20% more variance than what could be explained by our objective clutter estimates. Indeed, when we included both subjective and objective estimates as factors in a regression model, we found that subjective clutter alone could account for nearly all of the predicted variability in search RTs; the objective clutter estimate was not a significant contributor to the model.

Three conclusions follow directly from these observations. First, our observers were highly adept at estimating the clutter of these city scenes, even when the differences between scenes were incremental and small, and largely devoid of distinguishing semantic properties. Second, these subjective clutter estimates were highly predictive of search performance, a relationship reinforcing our opinion that clutter indeed does have the potential to serve as a meaningful measure of search load in the context of complex scenes. However, as a cautionary note, one should realize that even these impressive correlations between subjective clutter and search fall far short of the nearly perfect correlations often reported between set size and manual RTs using simpler stimuli (e.g., Treisman & Gelade, 1980). Third, subjective clutter estimates largely include the visual properties captured by objective clutter estimates, at least for the objective methods used in this study. On the one hand, this is informative and encouraging. It means that observers could incorporate fairly low-level visual attributes into their high-level clutter judgments, an ability that was not at all certain. It also means that subjective ratings can provide a sort of ground truth for objective clutter estimates, a benchmark against which models of clutter can be compared. On the other hand, the large differences in predictability between our subjective and objective measures means that current objective clutter estimates must be improved before they can be used as a valid surrogate measure of search set size. As for what these improvements should be, there are several possibilities. It might be that our subjective estimates included semantic factors and that these high-level factors helped predict search performance. If so, then objective clutter estimates need also to include semantic factors, a daunting task that is beyond the scope of existing clutter models. Alternatively, it may be that our subjective estimates were indeed dominated by visual factors but that existing objective models are not capturing these other visual dimensions that are important for predicting clutter. A final possibility is that our methods of computing objective clutter were inadequate. For example, it may be that the scale of the filter used in our edge detector was too large to capture much of the high-frequency clutter in the urban city, resulting in the disappointing correlations between search performance and edge density for that scene type. Sorting through these possibilities will be an important direction for future work.

In addition to manual measures, we also analyzed the eye movements made during search. Eye movement measures allow search behavior to be decomposed into smaller epochs, thereby enabling finer grained analyses (Zelinsky & Sheinberg, 1997). A clear example of this is the recent practice of segregating search behavior into guidance and verification components (e.g., Castelano et al., 2008; Malcolm & Henderson, 2009; Yang & Zelinsky, 2009). Doing this, we found a strong effect of clutter on search guidance; the path taken by gaze to the target was nearly four times less efficient for urban



cities compared to rural cities. Search path efficiency also decreased with increasing clutter as each type of city matured over time, although this relationship was found primarily in just the subjective clutter estimates. A similar relationship was found in the time needed to verify the target's presence after fixation. For all three scene types, increasing clutter resulted in longer target verification times. This is consistent with the suggestion that the time needed to discriminate a target from a background increases with target–background feature similarity (Boot et al., 2009; Neider & Zelinsky, 2006b; Wolfe et al., 2002).

These effects of clutter on search guidance and target verification can be broadly interpreted in terms of guided search theory (Wolfe, 1994a; Wolfe, Cave, & Franzel, 1989) and feature congestion, the suggestion that clutter effects arise when a scene's features begin to fill a given feature space (Rosenholtz et al., 2005, 2007). As visual clutter increases, so does the probability of non-target features matching the target, resulting in weakened guidance and a less direct search path to the target. Note that this can also be interpreted as an effect of target–distractor similarity; as clutter increases so does the variety of buildings appearing in a scene, thereby increasing the likelihood that one of these buildings will look like the target. A similar interpretation holds for target verification. As clutter increases, so does the probability of local background features matching the target, making the task of discriminating the target from the background more difficult. This potential for target–background discrimination to be affected by lower level visual processes might also explain why our objective clutter measures played a larger role in the case of target verification.

By analyzing the amplitude of the saccade used to bring gaze to the target, we also estimated how far in the visual periphery observers were able to detect a target and how this distance varied with clutter. This describes a sort of middle ground between guidance and target verification, reflecting the very strong guidance signal that presumably mediates the gaze movement that aligns fixation with a target. Although Henderson et al. (2009) found that saccade amplitudes were generally unaffected by clutter, our analysis of final saccade amplitudes suggested otherwise. Final saccade amplitudes were longest for the uncluttered rural scenes, shorter for the suburban scenes, and shorter still for the highly cluttered urban scenes. This effect, however, was limited to subjective clutter estimates and between-city clutter comparisons and must, therefore, be interpreted with caution. Still, this relationship suggests that clutter may affect the size of the area surrounding fixation over which an observer can extract high-quality information about the target, also known as the *useful field of view*, the *functional visual field*, or the *conspicuity area* (Ball, Beard, Roenker, Miller, & Griggs, 1988; Bouma, 1978; Engel, 1971, 1977; Mackworth, 1965). We speculate that this clutter effect may be caused by the same feature confusability between the local background and target that we believe resulted in the observed effect of

clutter on target verification; given that the probability of a confusable feature appearing near to the target increases with visual clutter, as clutter decreases, targets can be detected farther out in the visual periphery.

In conclusion, we found evidence for a relationship between visual clutter and a variety of search behaviors using evolving scenes that more closely approximated the incremental changes characterizing a standard set size manipulation. However, these relationships were strongest for our subjective estimates of clutter. Although our subjective and objective clutter estimates correlated highly with each other, objective clutter was a generally poorer predictor of search behavior, and the behavioral variability captured by these estimates could be explained almost entirely in terms of subjective clutter. This is a cause for concern among researchers seeking to use objective clutter as a surrogate for set size and raises the question of what this measure is missing that the subjective measure is not. Most obviously, objective measures would fail to capture semantic contributions to clutter. Whereas objective clutter might plateau early during a city's evolution, with each new structure adding edges but also erasing those from preexisting structures (a dynamic that might explain the generally lower correlations found between objective clutter and search performance in our urban scenes), a qualitatively different dynamic would likely exist for a city's semantic evolution. Schools and hospitals and skyscrapers may replace a conceptually homogeneous block of houses in a suburban city or a textured field in a rural scene. We believe that it is this *conceptual congestion* that is responsible for the more pronounced relationship between subjective clutter and search. The degree of conceptual congestion in a scene is likely related to type versus token distinctions in perception (e.g., Kanwisher, 1987, 1991). If a suburban neighborhood of 30 houses is replaced with a school and a park and an office building, in some sense there was a three-fold increase in the number of conceptual units; where once there were several tokens of a single house type, there are now three different types of structures. We speculate that it is the number of types of objects in a scene that drive clutter effects, not the number of tokens. Future work will explore this type/token distinction in the context of realistic scenes so as to better understand the role of conceptual congestion in real-world search.

## Acknowledgments

This work was supported by a Beckman Institute Postdoctoral Fellowship to M.B.N. and NIH Grant R01-MH063748 to G.J.Z. We would like to thank Chris Dickinson, Xin Chen, Hyejin Yang, Joe Schmidt, and all the other members of the Eye Cog Lab, especially Samantha Schmidt who was responsible for much of the stimulus creation and data collection.

Commercial relationships: none.

Corresponding author: Mark B. Neider.

Email: mark.neider@ucf.edu.

Address: Department of Psychology, University of Central Florida, 4000 Central Florida Blvd., Orlando, FL 32816-1390, USA.

## Footnotes

<sup>1</sup>Our implementation of the Canny edge detector used a sigma of 3, and no attempt was made to explore the range of this parameter to find values optimized for the city scenes used in this experiment. It is therefore possible that the use of different filter scales might have resulted in different estimates of edge density.

<sup>2</sup>Given that subjective clutter was rated separately for each scene type, one might argue that raters used different subjective clutter scales for the rural, suburban, and urban cities and that this compromises between-scene comparisons. While this cannot be ruled out, our data argue against this possibility. The use of different scales would serve to normalize for clutter differences between scenes, thereby discouraging the expression of clutter effects. The fact that we found large and significant differences in subjective clutter between scenes is more consistent with the use of a single subjective clutter scale.

<sup>3</sup>In general, correlations with log(RT) were generally higher than those with RT for both subjective clutter ratings and edge density, thereby largely preserving the relative differences between the two types of clutter estimates.

## References

- Baddeley, R. J., & Tatler, B. W. (2006). High frequency edge (but not contrast) predict where we fixate: A Bayesian system identification analysis. *Vision Research*, *46*, 2824–2833.
- Ball, K., Beard, B. L., Roenker, D. L., Miller, R. L., & Griggs, D. S. (1988). Age and visual search: Expanding the useful field of view. *Journal of the American Optometric Association*, *5*, 2210–2219.
- Beck, M. R., Lohrenz, M. C., & Trafton, J. G. (2010). Measuring search efficiency in complex visual search tasks: Global and local clutter. *Journal of Experimental Psychology: Applied*, *16*, 238–250.
- Biederman, I., Blicke, T. W., Teitelbaum, R. C., Klatsky, G. J., & Mezzanotte, R. J. (1988). Object identification in nonscene displays. *Journal of Experimental Psychology: Human Learning, Memory, and Cognition*, *14*, 456–467.
- Biederman, I., Glass, A. L., & Stacy, E. W. (1973). Searching for objects in real-world scenes. *Journal of Experimental Psychology*, *97*, 22–27.
- Boot, W. R., Neider, M. B., & Kramer, A. F. (2009). Training and transfer in search for camouflaged real-world targets. *Attention, Perception, & Psychophysics*, *71*, 950–963.
- Bouma, H. (1978). Visual search and reading: Eye movements and functional visual field: A tutorial review. In J. Requin (Ed.), *Attention and performance, VII* (pp. 115–146). Hillsdale, NJ: Erlbaum.
- Bravo, M. J., & Farid, H. (2004). Search for a category target in clutter. *Perception*, *33*, 643–652.
- Bravo, M. J., & Farid, H. (2006). Object recognition in dense clutter. *Perception & Psychophysics*, *68*, 911–918.
- Bravo, M. J., & Farid, H. (2008). A scale invariant measure of clutter. *Journal of Vision*, *8*(1):23, 1–9, <http://www.journalofvision.org/content/8/1/23>, doi:10.1167/8.1.23. [PubMed] [Article]
- Brockmole, J. R., & Henderson, J. M. (2006a). Recognition and attention guidance during contextual cueing in real-world scenes: Evidence from eye movements. *Quarterly Journal of Experimental Psychology*, *59*, 1177–1187.
- Brockmole, J. R., & Henderson, J. M. (2006b). Using real-world scenes as contextual cues during search. *Visual Cognition*, *13*, 99–108.
- Canny, J. (1986). A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *8*, 679–714.
- Castelano, M. S., & Henderson, J. M. (2007). Initial scene representations facilitate eye movement guidance in visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *33*, 753–763.
- Castelano, M. S., Pollatsek, A., & Cave, K. R. (2008). Typicality aids search for an unspecified target, but only in identification and not in attentional guidance. *Psychonomic Bulletin & Review*, *15*, 795–801.
- Chen, X., & Zelinsky, G. J. (2006). Real-world visual search is dominated by top-down guidance. *Vision Research*, *46*, 4118–4133.
- Craft, E., Schutze, H., Niebur, E., & von der Heydt, R. (2007). A neural model of figure-ground organization. *Journal of Neurophysiology*, *97*, 4310–4326.
- Driver, J., Davis, G., Russell, C., Turatto, M., & Freeman, E. (2001). Segmentation, attention and phenomenal visual objects. *Cognition*, *80*, 61–95.
- Duncan, J., & Humphreys, G. (1989). Visual search and stimulus similarity. *Psychological Review*, *96*, 433–458.

- 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.
- Eckstein, M. P., Drescher, B., & Shimozaki, S. S. (2006). Attentional cues in real scenes, saccadic targeting and Bayesian priors. *Psychological Science*, *17*, 973–980.
- Ehinger, K. A., Hidalgo-Sotelo, B., Torralba, A., & Oliva, A. (2009). Modelling search for people in 900 scenes: A combined source model of eye guidance. *Visual Cognition*, *17*, 945–978.
- Engel, F. (1971). Visual conspicuity, directed attention and retinal locus. *Vision Research*, *11*, 563–576.
- Engel, F. (1977). Visual conspicuity, visual search and fixation tendencies of the eye. *Vision Research*, *17*, 95–108.
- Foulsham, T., & Underwood, G. (2007). How does the purpose of inspection influence the potency of visual saliency in scene perception? *Perception*, *36*, 1123–1138.
- Greene, M. R., & Oliva, A. (2009). Recognition of natural scenes from global properties: Seeing the forest without representing the trees. *Cognitive Psychology*, *58*, 137–179.
- Henderson, J. M. (2003). Human gaze control in real-world scene perception. *Trends in Cognitive Sciences*, *7*, 498–504.
- Henderson, J. M. (2007). Regarding scenes. *Current Directions in Psychological Science*, *16*, 219–222.
- Henderson, J. M., Chanceaux, M., & Smith, T. J. (2009). The influence of clutter on real-world scene search: Evidence from search efficiency and eye movements. *Journal of Vision*, *9*(1):32, 1–8, <http://www.journalofvision.org/content/9/1/32>, doi:10.1167/9.1.32. [PubMed] [Article]
- 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.
- Hwang, A. D., Higgins, E. C., & Pomplun, M. (2009). A model of top-down attentional control during visual search in complex scenes. *Journal of Vision*, *9*(5):25, 1–18, <http://www.journalofvision.org/content/9/5/25>, doi:10.1167/9.5.25. [PubMed] [Article]
- Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shift of visual attention. *Vision Research*, *40*, 1489–1506.
- Kanan, C., Tong, M. H., Zhang, L., & Cottrell, G. W. (2009). SUN: Top-down saliency using natural statistics. *Visual Cognition*, *17*, 979–1003.
- Kanwisher, N. (1987). Repetition blindness: Type recognition without token individuation, *Cognition*, *27*, 117–143.
- Kanwisher, N. (1991). Repetition blindness and illusory conjunctions: Errors in binding visual types with visual tokens. *Journal of Experimental Psychology: Human Perception and Performance*, *17*, 404–421.
- Krendel, E. S., & Wodinsky, J. (1960). Search in an unstructured field. *Journal of the Optical Society of America*, *50*, 562–568.
- Lohrenz, M. C., Trafton, J. G., Beck, M. R., & Gendron, M. L. (2009). A model of clutter for complex, multivariate, geospatial displays. *Human Factors*, *51*, 90–101.
- Mackworth, N. H. (1965). Visual noise causes tunnel vision. *Psychonomic Science*, *3*, 67–68.
- Mackworth, N. H., & Morandi, A. J. (1967). The gaze selects informative details within pictures. *Perception & Psychophysics*, *2*, 547–552.
- Malcolm, G. L., & Henderson, J. M. (2009). The effects of target template specificity on visual search in real-world scenes: Evidence from eye movements. *Journal of Vision*, *9*(11):8, 1–13, <http://www.journalofvision.org/content/9/11/8>, doi:10.1167/9.11.8. [PubMed] [Article]
- Navalpakkam, V., & Itti, L. (2005). Modeling the influence of task on attention. *Vision Research*, *45*, 205–231.
- Neider, M. B., Boot, W. R., & Kramer, A. F. (2010). Visual search for real world targets under conditions of high target–background similarity: Exploring training and transfer of training in older adults. *Acta Psychologica*, *134*, 29–39.
- Neider, M. B., & Zelinsky, G. J. (2006a). Scene context guides eye movements during visual search. *Vision Research*, *46*, 614–621.
- Neider, M. B., & Zelinsky, G. J. (2006b). Searching for camouflaged targets: Effects of target–background similarity on visual search. *Vision Research*, *46*, 2217–2235.
- Neider, M. B., & Zelinsky, G. J. (2008). Exploring set size effects in realistic scenes: Identifying the objects of search. *Visual Cognition*, *16*, 1–10.
- Neider, M. B., & Zelinsky, G. J. (2010). Exploring the perceptual causes of search set-size effects in complex scenes. *Perception*, *39*, 780–794.
- Newell, F. N., Brown, V., & Findlay, J. M. (2004). Is object search mediated by object-based or image-based representations. *Spatial Vision*, *17*, 511–541.
- Oliva, A., & Torralba, A. (2006). Building the gist of a scene: The role of global image features in recognition. *Progress in Brain Research: Visual Perception*, *155*, 23–36.



- Oliva, A., & Torralba, A. (2007). The role of context in object recognition. *Trends in Cognitive Sciences*, *11*, 520–527.
- Oliva, A., Wolfe, J. M., & Arsenio, H. (2004). Panoramic search: The interaction of memory and vision in search through a familiar scene. *Journal of Experimental Psychology: Human Perception and Performance*, *30*, 1132–1146.
- Palmer, J., Verghese, P., & Pavel, M. (2000). The psychophysics of visual search. *Vision Research*, *40*, 1227–1268.
- Parkhurst, D. J., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual selective attention. *Vision Research*, *42*, 107–123.
- Pomplun, M. (2006). Saccadic selectivity in complex visual search displays. *Vision Research*, *46*, 1886–1900.
- Pomplun, M. (2007). Advancing area activation towards a general model of eye movements in visual search. In W. D. Gray (Ed.), *Integrated models of cognitive systems* (pp. 120–131). New York: Oxford University Press.
- Rao, R. P. N., Zelinsky, G. J., Hayhoe, M. M., & Ballard, D. H. (2002). Eye movements in iconic visual search. *Vision Research*, *42*, 1447–1463.
- Rosenholtz, R. (2001). Visual search for orientation among heterogeneous distractors: Experimental results and implications for signal-detection theory models of search. *Journal of Experimental Psychology: Human Perception and Performance*, *27*, 985–999.
- Rosenholtz, R., Li, Y., Mansfield, J., & Jin, Z. (2005). Feature congestion: A measure of display clutter. *SIGCHI 2005*, 761–770.
- Rosenholtz, R., Li, Y., & Nakano, L. (2007). Measuring visual clutter. *Journal of Vision*, *7*(2):17, 1–22, <http://www.journalofvision.org/content/7/2/17>, doi:10.1167/7.2.17. [PubMed] [Article]
- Schmidt, J., & Zelinsky, G. J. (2009). Search guidance is proportional to the categorical specificity of a target cue. *Quarterly Journal of Experimental Psychology*, *62*, 1904–1914.
- Tatler, B. W. (2009). Current understanding of eye guidance. *Visual Cognition*, *17*, 777–789.
- 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*, 766–786.
- Treisman, A., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*, 97–136.
- van den Berg, R., Cornelissen, F. W., & Roerdink, J. B. T. M. (2009). A crowding model of visual clutter. *Journal of Vision*, *9*(4):24, 1–11, <http://www.journalofvision.org/content/9/4/24>, doi:10.1167/9.4.24. [PubMed] [Article]
- Verghese, P. (2001). Visual search and attention: A signal detection theory approach. *Neuron*, *31*, 523–535.
- Võ, M. L.-H., & Henderson, J. M. (2010). Does gravity matter? Effects of semantic and syntactic inconsistencies on the allocation during scene perception. *Journal of Vision*, *9*(3):24, 1–15, <http://www.journalofvision.org/content/9/3/24>, doi:10.1167/9.3.24. [PubMed] [Article]
- Wertheimer, M. (1923). Untersuchungen zur Lehre von der Gestalt, II [Laws of organization in perceptual forms]. *Psychologische Forschung*, *4*, 301–350. Excerpts translated and reprinted in W. D. Ellis (Ed.) (1939). *A source book of Gestalt psychology* (pp. 71–88). New York: Harcourt, Brace and Co.
- Wolfe, J. M. (1994a). Guided search 2.0: A revised model of visual search. *Psychonomic Bulletin and Review*, *1*, 202–238.
- Wolfe, J. M. (1994b). Visual search in continuous, naturalistic scenes. *Vision Research*, *34*, 1187–1195.
- 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 1,000,000 trials tell us about visual search? *Psychological Science*, *9*, 33–39.
- Wolfe, J. M., Cave, K., & Franzel, S. (1989). Guided search: An alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *15*, 419–433.
- 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.
- Wolfe, J. M., Võ, M. L.-H., Evans, K. K., & Greene, M. R. (2011). Visual search in scenes involves selective and nonselective pathways. *Trends in Cognitive Sciences*, *15*, 77–84.
- Yang, H., & Zelinsky, G. J. (2009). Visual search is guided to categorically-defined targets. *Vision Research*, *49*, 2095–2103.
- Zelinsky, G. J. (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. J. (2001). Eye movements during change detection: Implications for search constraints, memory limitations, and scanning strategies. *Perception & Psychophysics*, *63*, 209–225.
- Zelinsky, G. J. (2008). A theory of eye movements during target acquisition. *Psychological Review*, *115*, 787–835.



Zelinsky, G. J., 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.

Zelinsky, G. J., & Schmidt, J. (2009). An effect of referential scene constraint on search implies scene segmentation. *Visual Cognition*, 17, 1004–1028.

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.