# Attributes of Subtle Cues for Facilitating Visual Search in Augmented Reality

Weiquan Lu, Henry Been-Lirn Duh, Steven Feiner, and Qi Zhao

**Abstract**—Goal-oriented visual search is performed when a person intentionally seeks a target in the visual environment. In augmented reality (AR) environments, visual search can be facilitated by augmenting virtual cues in the person's field of view. Traditional use of explicit AR cues can potentially degrade visual search performance due to the creation of distortions in the scene. An alternative to explicit cueing, known as subtle cueing, has been proposed as a clutter-neutral method to enhance visual search in video-see-through AR. However, the effects of subtle cueing are still not well understood, and more research is required to determine the optimal methods of applying subtle cueing in AR. We performed two experiments to investigate the variables of scene clutter, subtle cue opacity, size, and shape on visual search performance. We introduce a novel method of experimentally manipulating the scene clutter variable in a natural scene while controlling for other variables. The findings provide supporting evidence for the subtlety of the cue, and show that the clutter conditions of the scene can be used both as a global classifier, as well as a local performance measure.

**Index Terms**—Multimedia information systems—artificial, augmented, and virtual realities, user/machine systems: human factors

## 1 Introduction

Goal-oriented visual search is an action performed whenever a person seeks a target in the visual environment [1], [2]. In video-see-through augmented reality (AR) environments, AR visual cues can be used to direct the visual spotlight of a user to facilitate rapid visual search [3]. Traditional methods of visual cueing involve the augmentation of explicit virtual cues in the scene [4], [5], [6], [7], [8], [9], [10], [11]. Such use of explicit visual cues can increase the visual clutter in the scene, which may lead to degraded visual search performance [12].

Beyond the performance consideration, there exist situations where explicit cueing may be undesirable for the task. Examples of these are provided by previous work [3], [13], in which the attention capture capabilities of explicit cues conflict with the requirement for the user to maintain focused attention on a specific task at hand (which is different from the task facilitated by the explicit cues), and secondary nonintrusive means of attention redirection are required.

Alternatives to explicit cueing have been proposed [3], [13], [14], [15], [16]. These methods entail a more subtle approach to visual cueing. In particular, the terminology in [3] specifies a lightweight, artifact-based method of subtle cueing in AR that directs the user's visual spotlight while being almost imperceptible (as shown in Fig. 1). This results in the redeployment of visual attention without a significant change in visual clutter of the scene.

However, the concept of subtle cueing is still in its infancy. Rather than using heuristics, subtle cueing should be studied in a principled manner for the mechanism to be applied effectively in live AR applications. To help address this need, in this paper, we seek to investigate the attributes of subtle cueing, focusing on the parameters pertinent to the development of a subtle-cueing-based video-see-through AR system. By conducting two experimental studies, and comparing our results with those from previous studies in subtle cueing, we seek to form a better understanding of subtle cueing, and apply it in a more principled manner to real-world use cases.

The paper is organized as follows: First, Section 2 describes the related work in subtle cueing. Next, Section 3 details the methodology and describes two experiments we conducted. Section 4 presents the results and findings of the first experiment, and discusses how subtle cueing varies across a range of visual clutter conditions. Section 5 then discusses the results and findings of the second experiment, which investigates the attributes of cue shape and cue size in subtle cueing. Section 6 discusses our conclusions and implications of our research. Finally, Section 7 discusses the limitations of our approach and possible future work.

## 2 Related Work and Motivation

According to traditional literature, visual search consists of two components, namely the conspicuity of a target, and the expectancies associated with the target [2]. The conspicuity refers to how much the target stands out from the background, and is the basis on which the concept of visual saliency is founded [17], [18], [19]. The expectancies refer to a subject's expectation of where a target should be,

- *W. Lu, H. B.-L. Duh, and Q. Zhao are with the Interactive Digital Media Institute, National University of Singapore, #02-01-01 I3 Building, 21 Heng Mui Keng Terrace, Singapore 119613. E-mail: {lu.weiquan, eledbl, eleqiz}@nus.edu.sg.*
- *S. Feiner is with Columbia University, Department of Computer Science, 500 W. 120th St., 450 CS Building, New York, NY 10027.E-mail: feiner@cs.columbia.edu.*

*Manuscript received 23 Feb. 2013; revised 24 Aug. 2013; accepted 26 Sept. 2013; published online 3 Oct. 2013.*
*Recommended for acceptance by D. Bowman.*
*For information on obtaining reprints of this article, please send e-mail to: tvcg@computer.org, and reference IEEECS Log Number TVCG-2013-02-0050.*
*Digital Object Identifier no. 10.1109/TVCG.2013.241.*

1077-2626/14/$31.00 © 2014 IEEE Published by the IEEE Computer Society


Fig. 1. Visual comparison between explicit, subtle, and no cues.

what it should look like, and how it should behave, based on prior knowledge.

In effect, the two components of visual search exemplify the two principles of human attention: conspicuity can be said to be an example of bottom-up (stimulus-based) attention and the concept of expectancies is an example of top-down (experience-based) attention [17]. Bottom-up attention is dependent on the characteristics of the stimulus (such as color and motion) that elicit instinctual human responses and can be difficult to suppress. Top-down attention is influenced by human factors such as preknowledge, expectations, and goals. From this point of view, it stands to reason that one could either increase the target's conspicuity, or influence a person's expectancies of the target, to facilitate visual search.

However, this matter is complicated by the requirements of subtle cueing, in which the cueing mechanism has to be just barely noticeable to the user, yet still possess a significant cueing effect [3]. Hence, the traditional saliency-based approaches of evaluating human attention in a natural outdoor scene may not function well, especially since subtle cueing is meant to function in goal-oriented visual search scenarios, which traditional saliency-based methods were not designed to tackle [17]. A different approach to quantifying visual search performance may be required.

One such approach is presented by Rosenholtz et al. [12], using a measure of feature congestion (FC). FC is a method of calculating the amount of clutter of a scene. In turn, the clutter value has been shown to correlate well with general visual search performance. The reason why FC appears to model goal-oriented visual search performance well is because of the definition of the FC calculation. The premise of FC is that the visual system has an interest in detecting "unusual" items, and the less "unusual" an item is, the less probability that it will be noticed.

By analogy, when someone tries to add an object to a scene, it will more likely to be noticed in a less cluttered scene, than if it were added to a more cluttered scene, because in a more cluttered scene, the object has a lower chance of being "unusual." This is a departure from traditional saliency models that try to determine how much something "stands out" from the scene, since the amount to which something "stands out" also depends on its relative task importance in the scene based on expectancies, whereas FC simply assesses the degree of clutter as a state of the scene, without making assumptions about object importance.

In a sense, FC is a measure of the difficulty in reliably drawing attention to a newly added item in a scene, with the difficulty increasing as the amount of visual clutter increases. Simply speaking, FC allows the determination of the state of the scene, from which general visual search performance can be inferred. In their paper, Rosenholtz et al. studied the application of the FC measure in a repeated search task involving still images of geographic maps, to determine the contrast threshold between the cue and the context, for a wide range of FC between 2 and 12. The authors found that the results correlated well with the findings in previous visual clutter studies [20], [21], while extending the domain of visual search research beyond discrete, highly controlled laboratory-based stimuli, into the domain of continuous scenes. While their work was comprehensive, Rosenholtz et al. did not use dynamic or video scenes in their studies, and focused primarily on the validation of the FC measure using explicit cues, rather than on investigating cue characteristics of subtle cues.

Lu et al. [3] provide an overview of previous work on subtle cueing. In essence, subtle cueing is a clutter-neutral alternative to traditional explicit cueing. Lu et al. started from a simple premise: given a fixed target shape (a virtual cross superimposed on an outdoor background), they tried to determine the minimum amount of local contrast needed to allow the target to be found faster in a visual search task. In a series of experiments, the authors manipulated the local contrast of the area surrounding the target, by adjusting the opacity of a white square image (layered in between the target and the background). This white square acted as the subtle cue. Both still images and video of outdoor scenes were used, and the clutter content of the scenes were controlled by evaluating their FC [12] values. Only scenes that had FC between 5 and 6 were used.

From the results of the experiments, Lu et al. concluded that contrast-based subtle cueing was feasible for improving visual search performance, and found that there was a significant effect when the opacity (alpha channel value) of the subtle cue was between 0.1 and 0.3. This allowed the authors to determine the minimum contrast of cues required for subtle cueing to be effective, while still being almost imperceptible. However, their study was only conducted within a narrow range of clutter conditions, and did not evaluate other cue attributes such as cue size and cue shape.

The specific mechanism for subtle cueing in previous work [3] was based on contrast manipulation, as suggested by human vision studies of the attributes that have the potential to influence the deployment of visual attention [22], [23]. The effects of these contrast modulations were investigated in the context of visual search performance, which could be measured in relation to the amount of visual clutter in the scene.

Related to the concept of subtle cueing is the work done by Veas et al. [13] on using a saliency modulation technique (SMT) [16] to influence the deployment of visual attention. SMT is a method of image enhancement that uses saliency map techniques (as opposed to visual clutter methods) to modulate the contrast of a video scene image on a per-pixel basis. The purpose was to visually emphasize certain objects in the environment. The subtle modulations were designed to be perceptibly invisible to the observer, while still having significant influence on the observer's attention spotlight. Veas et al. focused on passively increasing the chances that visual attention would be drawn to specified real objects in the physical world, as opposed to enhancing active visual search for virtual objects in AR. Hence, Veas et al. applied the SMT method to enhance undirected viewing of a scene, instead of providing subtle cueing for directed search tasks.

Thus, while subtle cueing has been shown to be a feasible alternative to explicit cueing, it is necessary to investigate the attributes of subtle cueing required for the development of a subtle-cueing based AR system.

## 3 Methodology and Protocol

Our methodology, like that of Lu et al. [3], consists of two experiments designed in the fashion of typical human vision studies [12], [20], [21]—an observer searches for a virtual target, amidst a video background of an outdoor scene, simulating a video-see-through AR display. Details of the protocol will be described later in the paper. We also used the same contrast-based mechanism of subtle cueing as Lu et al. in a repeated search task.

We investigate three research questions not examined in prior studies:

- RQ1: Given the effect of opacity level of subtle cues on visual search performance, is there an interaction effect between opacity levels and scene clutter?
- RQ2: Within a scene, can FC be used both as a global scene classifier, as well as a local performance measure?
- RQ3: Is there empirical evidence to suggest that subjects were unable to notice the defining features of the cue, thereby providing evidence for that the cue is subtle and barely noticeable?

### 3.1 Manipulating Clutter in an Unprepared Scene

Previous work did not provide a method for manipulating the clutter content of a natural outdoor video scene in a controlled manner. Therefore, we had to devise a method that kept the context of the video scene relatively controlled, while still allowing FC to vary in a predictable manner, to allow for repeatability and validity of the experiment conditions.

We achieved this by using a Microsoft Cinema HD webcam in a stationary position, overlooking an outdoor scene. Scene footage was captured using the webcam at 30 frames-per-second, over a course of 30 hours. The footage was then analyzed frame by frame (using the FC method of Rosenholtz et al. [12]) to generate a visual clutter profile of the scene across time. A 10-second sliding window was then applied over the entire recoding, which allowed a 10-second video clip to be selected for desired characteristics such as scene content, average brightness, and clutter conditions. The FC across the entire recording was within the range 1.1-6.8. We analyzed the scene clutter profile to uncover the reasons for this variability (refer to Fig. 2).


Fig. 2. Comparing the appearance of the same object in images $A$ and $B$. Areas in the key image $S$ are tinted red in proportion to the difference between the clutter profiles of images $A$ and $B$ at those areas. The difference is a result of shadows and reflections.

While it is obvious that visibility of objects would affect the clutter profile of the scene, we focused on uncovering factors that were more subtle and difficult to notice by the naked eye. To do this, we took two images, shown in Fig. 2, that were close to each other in terms of their clutter profiles: Image $A$ had an FC value of 4.5, and Image $B$ had an FC value of 5.0. We processed them using the following steps:

1. Clutter maps (as defined in [12]) for each image were generated.
2. Subtraction of the clutter maps produced a single grayscale image
3. The grayscale image was inverted to allow for easier visualization
4. The inverted image was enhanced by red false-coloring.
5. The red false-color image was superimposed on Image $A$ to for visual analysis, producing Image $S$.

### 3.2 Uncovering Reasons for Changes in Scene Clutter

Image $S$ allowed us to look for regions of FC difference. We excluded analysis of areas where objects (such as cars and branches) were prominently present in one image and were absent or had moved in the next image, since that would give us no additional insight beyond the obvious.

We found that besides the presence and absence of physical objects, the different environmental conditions throughout the day created situations in which the position of shadows and reflections changed as the day progressed. A sample of such a situation is shown in Fig. 2, which is illustrative of the larger phenomenon happening throughout the entire scene. It is interesting to note that these seemingly subtle and unpredictable changes, along with the more obvious presence and absence of physical objects, contributed significantly to the variation in the clutter conditions. Together with the varying lighting conditions that occurred throughout the course of the 30-hours

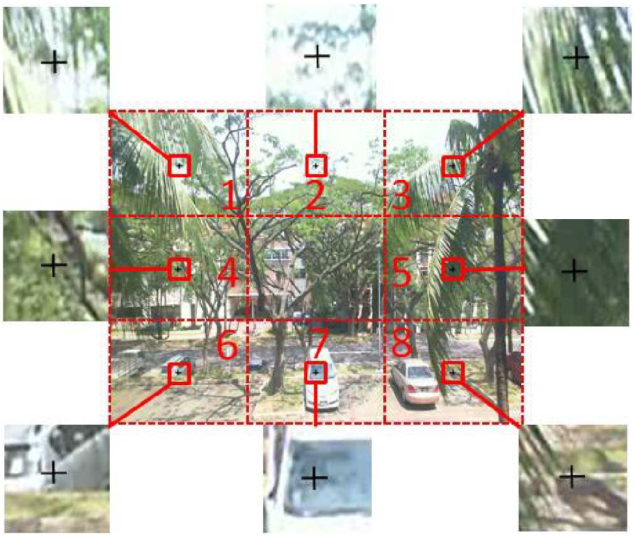


Fig. 3. Eight target positions (within eight local segments). Local segments will be used for standardized FC and luminance calculations for analysis. Red guidelines and labels are for illustration purposes.

recording, the FC variation could be characterized according to these factors.

For experimental purposes, only the range 4-6 FC (daytime content) was used in our study, as values outside of this range were no longer considered as having similar content characteristics (as indicated by participants in a pretest). Even so, this selected range allowed us to increase the range of FC values investigated to double the range as compared to Lu et al. [3]. The selected clips were then used as the background video upon which the subtle cue and target were superimposed.

### 3.3 Experiment Stimuli

In both experiments, subjects searched for a target, a black cross "+", in a video of an outdoor scene (as captured from the webcam). Such a target cross might be used in military AR applications, such as those mentioned in Livingston et al. [24]. The cross was $13 \times 13$ pixels, approximately subtending a visual angle of 0.36 degree from a reasonable viewing position and was embedded in a video scene of $1024 \times 768$ pixels, which subtended a visual angle of approximately 29.8 degree horizontal and 22.6 degree vertical.

For each experiment trial, the target location in the outdoor scene could be in any one of eight isoeccentric locations (see Fig. 3), at approximately 8.1 degree from the initial fixation point in the center of the image. These fixed positions were used to assist in standardizing the area from which local FC and contrast would be calculated, as opposed to a random location that would not allow equal comparison across trials. As will be shown in the results, there were no significant order effects. No target appeared at the center of the image, since that location would be the initial gaze starting point as mandated in the experiment protocol [3], [12]. A white mask and black fixation point were provided between trials to facilitate the initial gaze starting point. The video scene formed the background layer. A white square (with adjustable opacity), acting as

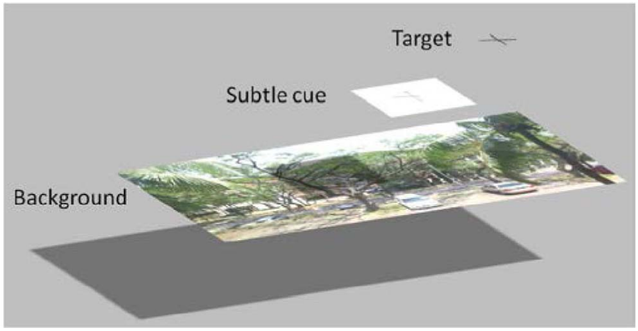


Fig. 4. Constructing the subtle cue by layering a white square in between the background and the target. The opacity of the white square can be varied to manipulate contrast.

the subtle cue, was superimposed on the background. On top of the subtle cue, we layered the target (refer to Fig. 4).

For all trials, target presence and location were counterbalanced to prevent learning and order effects. Hence, subjects could not predict where and when the target would be present, as in the previous work [3]. Specific details of experiment conditions are provided in the following sections.

### 3.4 Experiment Variables and Parameters

In both experiments, there were four independent variables and two dependent variables. The independent variables were:

- *Scene clutter*. By selecting video scenes with the desired profile, the scene clutter variable was manipulated by displaying the backgrounds (as ten-second video clips) that had the required global FC profiles.
- *Cue opacity*. The variable of cue opacity was implemented by varying the opacity (alpha channel value) of the subtle cue, within the range 0-1. For example, a value of 0 would make the square totally transparent, and a value of 1 would be totally opaque. See Fig. 1 for an illustration.
- *Cue size*. The cue size was determined as integral multiples of the target size in terms of visual angle. Hence, since the target was $13 \times 13$ pixel (approximately 0.36 degree from a reasonable viewing position), cue sizes of approximately $13 \times 13$, $26 \times 26$, $39 \times 39$, $52 \times 52$, and $65 \times 65$ pixels (corresponding, respectively, to approximately 0.36, 0.72, 1.08, 1.44, and 1.8 degrees from a reasonable viewing position), were used in these experiments.
- *Cue shape*. Three cue shapes were used ("square," "circle," "equilateral triangle"), with the target in the center of the cue. The area of the shapes was controlled based on the visual angle at a given cue size as shown above. The edges of the shapes were constrained to coincide with the visual angles. These were primary shapes [25] and were arbitrarily chosen for their difference in features between one another.

The dependent variables were:

- *Reaction time (RT)*. RT was measured as the time difference between the time recorded from the initial display of the stimulus, to the time the subject indicated the target being absent/present.

- *Error rate (ER)*. ER was measured by averaging the number of wrong responses over the total number of responses per subject.

In analyzing both experiments, we assumed $p < 0.05$ to be significant.

### 3.5 Experiment Protocol

Following Lu et al. [3], the following protocol was common to experiments $A$ and $B$. The test subject was brought into a darkened room and was seated in front of and facing a computer screen. The room was darkened to minimize light pollution from external sources. The display was a Philips Brilliance 240PW 24-inch monitor, using default factory settings, connected to a Windows PC.

Subjects were instructed to search for the target in the video background. The subjects were then to respond as quickly and as accurately as possible (by pressing a button on the keyboard) to indicate whether the target was present or absent. This experimental method has been used in previous visual search research [12]. However, the experimental method required justification for the 50 percent probability of chance that the subjects would respond randomly. We managed this issue through our protocol, by using rewards (prizes and recognition on an online leaderboard) to motivate the subject to perform within a 15 percent error rate, while maintaining an average of five seconds per trial. These conditions were designed to be challenging for the subject, and required conscious effort to meet. We only used subject data if they performed within the 15 percent error rate, which helped minimize the possibility of chance responses skewing the results.

At the start of each trial, a black fixation cross on a pure white background was displayed on the screen. After 500 ms, the stimulus replaced the fixation cross. A timer was started at this point, and counted until the test subject responded by pressing the left or right arrow on the keyboard, indicating that the target was absent or present, respectively. Once the key was pressed, the timer was stopped, and the elapsed time was recorded. The stimulus was then replaced by the fixation cross against the white background, and the next trial began.

As each experiment session lasted about 40 minutes, subjects were at risk of suffering from fatigue, frustration, and boredom, especially if they did not know their progress. Hence, after 10 percent of the trials per subject for each experiment, an interval screen was shown to the subject, including the average time taken per trial, the number of errors made, and a timer that counted from 0 to 30 seconds.

This interval screen served three functions. First, the subject was allowed to rest while it was displayed, thereby reducing fatigue. Second, it allowed subjects to know their progress, thereby managing their expectations and reducing frustration about when the experiment would finish. Third, it provided subjects with a sense of motivation to carry on with the experiment and perhaps win the prize, thereby preventing boredom from setting in.

We note here the concern that providing such feedback might bias the data and create an order effect, even though the method has been used in previous work [3], [12]. An analysis of the experiment data in the Section 4 shows that there are indeed no order effects.

Before each experiment, the subjects were briefed on the task, and were given an eyesight and Ishihara color test [26] to ensure that their vision was normal or corrected to normal. The subjects were also allowed a set of practice trials to familiarize themselves with the task. These practice trials could be repeated as many times as the subjects wanted, and they could ask the experimenter any question about the task at this time. Practice trials had different stimuli from the actual trials.

An important objective of these familiarization trials was to ensure that subjects were trained to use the left and right arrow keys for the correct responses. To allow subjects to learn and be acclimated to the input interface quickly, a conceptual mapping was taught to the subjects using the phrase "Left arrow for Left Out (as in the target was LEFT OUT of the image), and right arrow for Right There (as in the target is RIGHT THERE in the image)." The actual experiment was started only after subjects demonstrated full understanding of the response methods, as determined through an assessment by the experimenter.

### 3.6 Experiment $A$-Specific Stimuli

Twenty-seven university students (8 female, mean $\mathrm{age} = 26.5, \mathrm{SD} = 3.50$) participated in Experiment $A$. All subjects had normal or corrected-to-normal eyesight. Data from four subjects were rejected, as they failed to meet the required 15 percent error rate.

The goal of Experiment $A$ was to investigate the effect of different clutter conditions (FC values) on visual search performance, given different subtle cue opacity levels. Hence, the variables of cue size and cue shape were kept constant (cue size $\approx$ 0.72 degree, cue shape = "square"), and only the variables of scene clutter and cue opacity were adjusted. Five 10-second video clips were selected, each with a specific FC profile (4, 4.5, 5, 5.5, and 6). The cue opacity had three levels (0, 0.1, 0.2), as recommended in previous work [3]. In previous work, alpha value 0.1 was considered the minimum opacity level for subtle cueing to function, 0.2 was taken as a reference level, and value 0 was taken as the control since the cue was fully transparent. These levels were determined through pretesting, as in [3].

Therefore, the experiment was conducted with (5 video clips) $\times$ (2 target conditions: present or absent) $\times$ (8 target locations) $\times$ (3 opacity levels), for a total of 240 possible trials per subject. The trials were counterbalanced for all variables.

### 3.7 Experiment $B$-Specific Stimuli

Twenty-six university students (9 female, mean age = 27.6, $\mathrm{SD} = 3.44$ ) participated in Experiment $B$. All subjects had normal or corrected-to-normal eyesight. Data from three subjects were rejected as they failed to meet the required 15 percent error rate.

As a follow-up experiment to Experiment A, the goal of Experiment $B$ was to investigate the effect of cue size and cue shape on visual search performance within a given

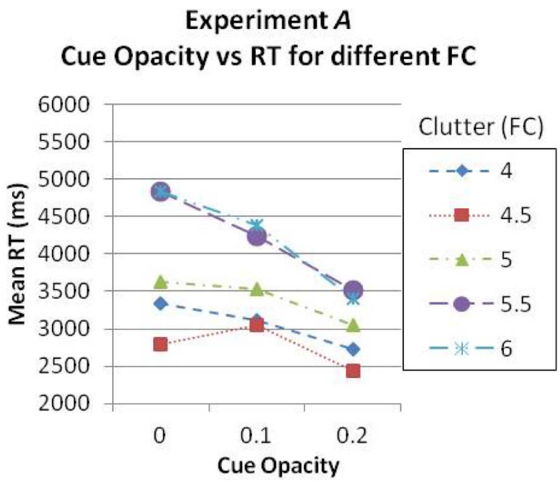


Fig. 5. Graph showing lack of interaction effect of FC and cue opacity on RT.

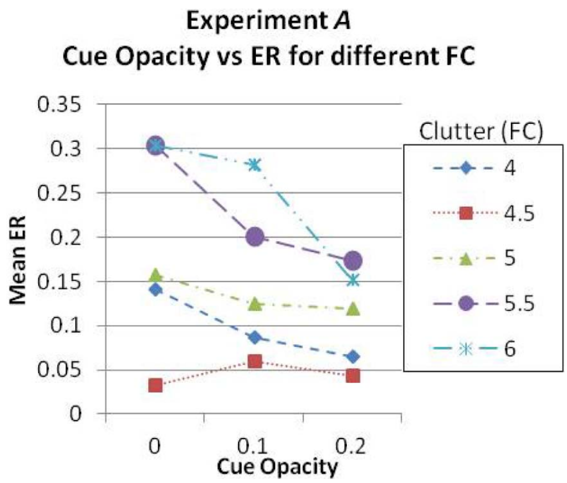


Fig. 6. Graph showing lack of an interaction effect of FC and cue opacity.

clutter condition and opacity level. Hence, the variables of clutter condition and cue opacity were kept constant ($FC = 6$, cue opacity $= 0.1$), and only the variables of cue size and cue shape were adjusted. $FC = 6$ was chosen as it had the most clutter among the tested clutter conditions, and cue opacity $= 0.1$ was selected because that it was the minimum level tested where subtle cueing functioned. Five cue sizes were investigated (corresponding to approximately 0.36, 0.72, 1.08, 1.44, and 1.8 degrees). Three cue shapes were investigated ("square," "circle," and "triangle"). The control was taken as the "square" cue shape at 0.716 degree, as used in Experiment $A$ and similar to previous work [3].

Therefore, the experiment was conducted with (1 video clip) $\times$ (2 target conditions: present or absent) $\times$ (8 target locations) $\times$ (5 cue sizes) $\times$ (3 cue shapes) for a total of 240 possible trials per subject. The trials were counterbalanced for all variables.

## 4 Results of Experiment $A$

Using repeated measures ANOVA and $t$-tests, we analyzed the results for the target-present cases.

### 4.1 Analysis of Global Scene Clutter Conditions

In terms of RT, the results showed a significant main effect ($F[2,44] = 64.285$, $p < 0.01$) across all three opacity levels (see Fig. 6). Using paired samples $t$-tests between opacity levels 0 and 0.1 ($t[22] = 2.847$, $p < 0.01$), as well as levels 0.1 and 0.2 ($t[22] = 8.406$, $p < 0.01$), the results were significant.

When analyzed from the perspective of the five scene clutter conditions, taking opacity level 0.1 as reference, there was a significant main effect of scene clutter ($F[4,88] = 95.315$, $p < 0.01$) across all five FC levels. Using paired samples $t$-tests between FC levels 4 and 4.5 ($t[22] = 0.303$, $p = 0.764$), 4.5 and 5 ($t[22] = -2.128$, $p < 0.05$), 5 and 5.5 ($t[22] = -3.766$, $p < 0.01$), and 5.5 and 6 ($t[22] = -0.558$, $p < 0.583$), mean differences between levels 4.5 and 5, as well as 5 and 5.5, were significant (see Fig. 6).

No significant interaction effect between scene clutter and cue opacity was found.

In terms of ER, the results showed a significant main effect ($F[2,44] = 11.961$, $p < 0.01$) across all three opacity levels. Using paired samples $t$-tests between opacity levels 0 and 0.1 ($t[22] = 2.148$, $p < 0.05$), as well as levels 0.1 and 0.2 ($t[22] = 2.940$, $p < 0.01$), the results were significant.

When analyzed from the perspective of the five scene clutter conditions, taking opacity level 0.1 as reference, there was a significant main effect of scene clutter ($F[4,88] = 37.823$, $p < 0.01$) across all five FC levels. Using paired samples $t$-tests between FC levels 4 and 4.5 ($t[22] = 1.000$, $p = 0.328$), 4.5 and 5 ($t[22] = -2.787$, $p < 0.05$), 5 and 5.5 ($t[22] = -2.440$, $p < 0.05$), as well as 5.5 and 6 ($t[22] = -1.875$, $p = 0.074$), mean differences between levels 4.5 and 5, as well as 5 and 5.5 were significant. No significant interaction effect between scene clutter and cue opacity was found (see Fig. 6).

### 4.2 Lack of Order Effects

To address concerns that the experiment protocol could have biased the results, a Pearson correlation analysis between the trial order and performance (RT and ER) was done. There was no strong correlation between trial order and performance (RT: $r = 0.105$, ER: $r = 0.016$).

### 4.3 Global verses Local Scene Clutter

To address RQ2, we needed to determine if opacity manipulations had the same effect on local scene clutter conditions as on global scene clutter. To do this, we divided the scene into nine equal segments, eight of which corresponded with the eight target locations at the center of the segments (refer to Fig. 3). By utilizing the same methods used to analyze the global scene, we applied these methods to the analysis of each individual segment. This allowed us to characterize the performance in each local segment.

In terms of RT, the results for all three opacity levels across each segment are as follows: Segment 1 ($F[2,44] = 0.418$, $p = 0.661$), Segment 2 ($F[2,44] = 8.316$, $p < 0.01$), Segment 3 ($F[2,44] = 24.758$, $p < 0.01$), Segment 4 ($F[2,44] = 3.701$, $p < 0.05$), Segment 5 ($F[2,44] = 42.970$, $p < 0.01$), Segment 6 ($F[2,44] = 5.424$, $p < 0.01$), Segment 7 ($F[2,44] = 3.556$, $p < 0.05$), and Segment 8 ($F[2,44] = 6.444$, $p < 0.01$). Refer to Fig. 7 for a detailed representation.

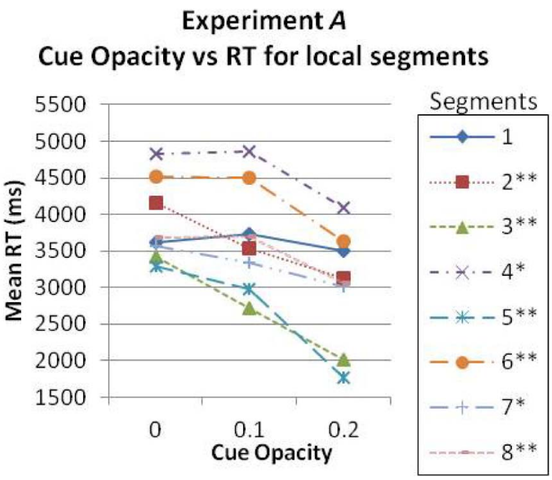


Fig. 7. Graph of RT in local segments. **denotes $p < 0.01$, *denotes $p < 0.05$.

In terms of ER, the results for all three opacity levels across each segment are as follows: Segment 1 ($F[2, 44] = 0.517$, $p = 0.600$), Segment 2 ($F[2, 44] = 6.292$, $p < 0.01$), Segment 3 ($F[2, 44] = 9.846$, $p < 0.01$), Segment 4 ($F[2, 44] = 0.272$, $p = 0.763$), Segment 5 ($F[2, 44] = 7.835$, $p < 0.01$), Segment 6 ($F[2, 44] = 5.642$, $p < 0.01$), Segment 7 ($F[2, 44] = 1.309$, $p = 0.280$), and Segment 8 ($F[2, 44] = 2.793$, $p = 0.109$). Refer to Fig. 8 for a detailed representation.

To assess the similarity in performance between global and local scene clutter, a Pearson correlation analysis was done between the performance data of global scene clutter, and the performance data of local scene clutter subdivided into individual segments. Table 1 shows the correlation matrix.

To assess the effect of cue opacity on contrast, the Michelson contrast [27] calculation was applied to compare the contrast between the three cue opacity conditions for each segment. Table 2 shows the results of the calculations. A Pearson correlation analysis between the significance of the results of cue opacity versus RT and ER in local segments (illustrated in Figs. 7 and 8) and delta values of the Michelson contrasts (Table 2) shows that there is substantial correlation between the results (RT: $r = -0.402$, ER: $r = -0.593$).

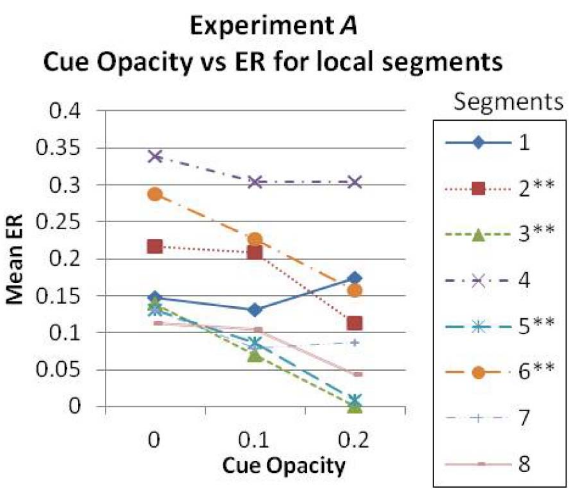


Fig. 8. Graph of ER in local segments. **denotes $p < 0.01$.

TABLE 1
Correlation of Performance between Global and Local Segments for Three Cue Opacity Conditions (0, 0.1, 0.2)

| | Pearson's *r* | | | | | |
|---|---|---|---|---|---|---|
| | RT | | | ER | | |
| Segment | 0 | 0.1 | 0.2 | 0 | 0.1 | 0.2 |
| 1 | .601 | .473 | .854 | .554 | .816 | .535 |
| 2 | .524 | .681 | .663 | .605 | .490 | .644 |
| 3 | .732 | .605 | .710 | .630 | .617 | .967 |
| 4 | .604 | .577 | .801 | .754 | .440 | .669 |
| 5 | .530 | .765 | .501 | .655 | .369 | .305 |
| 6 | .473 | .202 | .201 | .494 | .259 | .658 |
| 7 | .600 | .547 | .633 | .556 | .394 | .645 |
| 8 | .468 | .449 | .066 | .591 | .337 | .399 |

## 5 Results of Experiment $B$

Using repeated measures ANOVA and $t$-tests, we analyzed the results for the target-present cases.

In terms of RT, the results showed a significant main effect ($F[4, 88] = 4.723$, $p < 0.01$) across all five cue sizes. Using paired samples $t$-tests between cue sizes 0.36 and 0.72 degrees ($t[22] = 2.263$, $p < 0.05$), 0.72 and 1.08 degrees ($t[22] = 0.807$, $p = 0.428$), 1.08 and 1.44 degrees ($t[22] = 0.943$, $p = 0.356$), 1.44 and 1.8 degrees ($t[22] = -1.935$, $p = 0.066$), as well as 0.36 and 1.45 degrees ($t[22] = 3.687$, $p < 0.01$) mean differences between sizes 0.36 and 0.72 degrees, as well as between 0.36 and 1.44 degrees, were significant (Fig. 9).

When analyzed from the perspective of the three shape conditions, taking cue size 1.44 degree as the reference, there was no significant main effect of cue shape ($F[2, 44] = 0.390$, $p = 0.679$) across all three shapes.

In terms of ER, the results showed a significant main effect ($F[4, 88] = 5.063, p < 0.01$) across all five cue size levels. Using

TABLE 2
Michelson Contrast between Cue Opacity Conditions (0 versus 0.1, 0.1 versus 02)

| Segment | Michelson Contrast (%) | | Δ |
|---|---|---|---|
| | 0 vs 0.1 | 0.1 vs 0.2 | |
| 1 | 2.86 | 2.86 | 0 |
| 2 | 0.88 | 0.96 | 0.77 |
| 3 | 3.91 | 3.72 | 0.19 |
| 4 | 4.65 | 4.70 | 0.05 |
| 5 | 5.88 | 5.46 | 0.42 |
| 6 | 4.08 | 3.95 | 0.13 |
| 7 | 3.78 | 3.73 | 0.05 |
| 8 | 4.34 | 4.12 | 0.22 |

*Δ denotes the difference (absolute) between Michelson contrasts (0 versus 0.1) and (0.1 versus 0.2).*

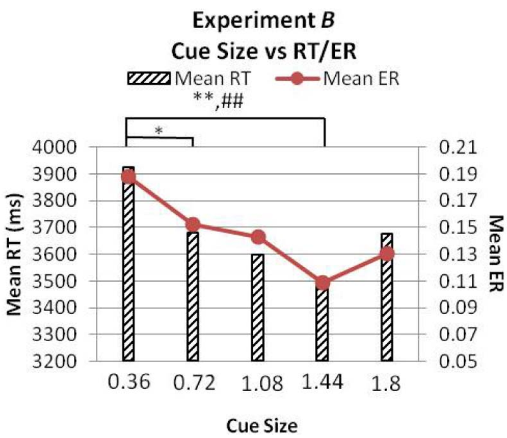


Fig. 9. Graph of cue size versus RT/ER. *denotes $p < 0.05$ for RT. ** denotes $p < 0.01$ for RT. ## denotes $p < 0.01$ for ER.

paired samples $t$-tests between cue sizes 0.36 and 0.72 degrees ($\mathrm{t}[22] = 1.950, p = 0.064$), 0.72 and 1.08 degrees ($\mathrm{t}[22] = 0.417$, $p = 0.681$), 1.08 and 1.44 degrees ($\mathrm{t}[22] = 2.060$, $p = 0.051$), 1.44 and 1.8 degrees ($\mathrm{t}[22] = -1.238, p = 0.229$), as well as 0.36 and 1.44 degrees ($\mathrm{t}[22] = 4.815$, $p < 0.01$), mean differences between sizes 0.36 and 1.44 degrees were significant (Fig. 9).

When analyzed from the perspective of the three shape conditions, taking cue size 1.44 degree as reference, there was no significant main effect of cue shape ($\mathrm{F}[2, 42] = 0.247$, $p = 0.782$) across all three shapes.

## 6 Discussion and Conclusions

Reviewing the results of Experiment $A$ to examine RQ1, there does not appear to be an interaction effect between scene clutter and opacity levels (as shown in Figs. 5 and 6). The results support the findings of previous work, in that an increase in FC resulted in decreased visual search performance [12]. The results also support the work of Lu et al. [3], in which the minimum opacity level at 0.1 is shown to be effective in improving performance.

To address RQ2, the correlation matrix in Table 1 provides evidence that, in general, the local trends mimic the global trends (Mean RT $r = 0.553$, SD RT $r = 0.189$, Mean ER $r = 0.558$, SD ER $r = 0.167$). This suggests that both local and global performance are proportional to the amount of clutter sampled in a given area of the scene. In turn, the findings support the argument that FC can be used both as a global attribute to classify the general visual search performance in the scene, as well as a local performance measure, depending on the scale in which it is measured. The results of the analysis done using Michelson contrasts suggest that contrast is indeed a key driver of performance, since the significance of the performance differences increases as the difference in Michelson contrast between different cue opacity conditions increases. This also suggests that there may be specific contrast thresholds that are related to performance, since even small changes in contrast can result in a difference between significance levels. It is interesting to note that a small minority of outliers exist, which could be due to specific (and perhaps isolated) characteristics within those samples. This provides motivation for future study.

The results of Experiment $B$ suggest that increasing cue size improves visual search performance (see Fig. 9), and there does not seem to be an optimum cue shape. The results regarding cue size are in line with expectations, as they support the results of previous work regarding size as an attribute that guides the deployment of attention [22]. Specifically, the larger the size, the higher the probability that the cue will be detected, in spite of its subtlety.

Addressing RQ3, the results regarding cue shape are interesting, because in previous work, shape as an attribute was considered as a probable candidate for attention guidance [22]. In subtle cueing, perhaps the reason why shape had no significant effect for the given cue size, was that, unlike explicit cues that are meant to be attention grabbing, the subtlety of the subtle cue prevented its features from being noticed. Hence, no matter the shape, the subtle cue should still function as prescribed by other attributes. This gives supporting evidence for the claim of cue subtlety.

Regarding the experimental method, the changes in clutter conditions were affected by scene visibility, object presence (or absence), reflections off shiny surfaces of scene objects, as well as shadows created by different lighting conditions throughout the day. Hence, it was possible to select for different clutter conditions, yet still keep the overall context of the scene similar, since reflections and shadows could change without the scene content actually changing significantly.

In conclusion, our current work has made contributions to the domain of subtle cueing in outdoor video scenes, with the potential of being used in video-see-through AR. First, we have introduced a method of manipulating the variable of clutter in unprepared outdoor scenes while keeping other variables such as context relatively constant. Second, our results have supported the findings of previous studies [3], [12], and found no interaction effect between clutter conditions and cue opacity. Third, we have found that both global and local performance can be characterized in the same way by the clutter conditions of the region sampled. Finally, our work has investigated the attributes of cue size and shape, and found supporting evidence for the subtlety of the cue at the given cue size and clutter condition.

## 7 Limitations and Future Work

There are many limitations to this work. First and foremost, our work has only focused on a stationary camera video-see-through system, and how subtle cueing functions in moving camera scenarios remains unknown. To investigate moving camera scenarios properly without the associated issues of motion sickness, future work will need to address the conditions of motion and orientation of the subtle cues against a moving background, preferably affording the user free movement. This suggestion involves further process and system complexity due to the need for accurate tracking and registration.

Second, this work is not yet applicable to conventional optical-see-through AR systems, since the user's perception of the physical world cannot be accessed in the same way as their perception of a video-see-through display [28]. While there is a possibility that the concepts could be applied to optical-see-through AR, this hypothesis needs to be tested in future studies.

Third, the experiments were conducted within a restricted range of visual clutter conditions ($4 < \mathrm{FC} < 6$). While many outdoor scenes fall within this range (as seen from the clutter profile of our captured footage), other environments may present clutter conditions outside of this range. However, as seen from our experiments, not only is the clutter condition important, but the other associated scene elements (such as context) have to be well controlled for experimental validity and reliability to be achieved.

Fourth, this work has investigated only a small subset of the attributes and parameters that could potentially be applied to subtle cueing. Other attributes such as luminance onset, closure, curvature, line termination, and expansion should be studied to provide subtle cue designers with a more complete set of tools to implement subtle cueing.

Fifth, this work has only investigated single target scenarios. Further work is required to determine how subtle cueing can be applied to multiple target scenarios.

Last but not least, while these experiments were conducted in a well-controlled laboratory setting, and even though the approach was justified in previous work [3], the performance of subtle cueing in outdoor AR is still an open question. Hence, we believe that it will be necessary to implement subtle cueing in a working portable prototype, so as to discover the practical advantages and limitations of subtle cueing in outdoor AR environments. Here, we note that the performance of the FC algorithm is still too slow for real-time operation (one frame per second). This technical challenge has to be overcome before a real-time prototype can be realized.

## Acknowledgments

This research was supported by the Singapore National Research Foundation under its International Research Center Keio-NUS CUTE Center @ Singapore Funding Initiative and administered by the IDM Program Office. Feiner was funded in part by US National Science Foundation (NSF) Grant IIS-0905569.

## References

[1] J.M. Wolfe, "Visual Search," *Current Biology,* vol. 20, no. 8, pp. R346-R349, Feb. 2010.

[2] C.D. Wickens, J.D. Lee, Y. Liu, and S.E. G. Becker, "Visual Search and Detection," *An Introduction to Human Factors Eng.,* second ed., L. Jewell, ed., pp. 78-90, Prentice Hall, 2004.

[3] W. Lu, H.B.-L. Duh, and S. Feiner, "Subtle Cueing for Visual Search in Augmented Reality," *Proc. IEEE Int'l Symp. Mixed and Augmented Reality (ISMAR '12),* pp. 161-166, 2012.

[4] L. Bonanni, C.-H. Lee, and T. Selker, "Attention-Based Design of Augmented Reality Interfaces," *Proc. Extended Abstracts on Human Factors in Computing Systems (CHI '05),* pp. 1228-1231, 2005.

[5] S.D. Peterson, M. Axholt, and S.R. Ellis, "Label Segregation by Remapping Stereoscopic Depth in Far-Field Augmented Reality," *Proc. Seventh IEEE/ACM Int'l Symp. Mixed and Augmented Reality (ISMAR '08),* pp. 143-152, Sept. 2008.

[6] S.D. Peterson, M. Axholt, M. Cooper, and S.R. Ellis, "Visual Clutter Management in Augmented Reality: Effects of Three Label Separation Methods on Spatial Judgments," *Proc. IEEE Symp. 3D User Interfaces (3DUI '09),* pp. 111-118, 2009.

[7] J. Wither, S. DiVerdi, and T. Höllerer, "Annotation in Outdoor Augmented Reality," *Computers & Graphics,* vol. 33, no. 6, pp. 679-689, Dec. 2009.

[8] M. Billinghurst, J. Bowskill, N. Dyer, and J. Morphett, "An Evaluation of Wearable Information Spaces," *Proc. Virtual Reality Ann. Int'l Symp. (VRAIS '98),* pp. 20-27, 1998.

[9] B. Schwerdtfeger and G. Klinker, "Supporting Order Picking with Augmented Reality," *Proc. Seventh IEEE/ACM Int'l Symp. Mixed and Augmented Reality (ISMAR '08),* pp. 91-94, 2008.

[10] F. Biocca, C. Owen, A. Tang, and C. Bohil, "Attention Issues in Spatial Information Systems: Directing Mobile Users' Visual Attention Using Augmented Reality," *J. Management Information Systems,* vol. 23, no. 4, pp. 163-184, May 2007.

[11] F. Biocca, A. Tang, and C. Owen, "Attention Funnel: Omnidirectional 3D Cursor for Mobile Augmented Reality Platforms," *Proc. SIGCHI Conf. Human Factors in Computing Systems (SIGCHI '06),* pp. 1115-1122, 2006.

[12] R. Rosenholtz, Y. Li, and L. Nakano, "Measuring Visual Clutter," *J. Vision,* vol. 7, no. 2, pp. 1-22, 2007.

[13] E. Veas, E. Mendez, S. Feiner, and D. Schmalstieg, "Directing Attention and Influencing Memory with Visual Saliency Modulation," *Proc. SIGCHI Conf. Human Factors in Computing Systems (CHI '11),* pp. 1471-1480, 2011.

[14] S. Su, F. Durand, and M. Agrawala, "De-Emphasis of Distracting Image Regions Using Texture Power Maps," *Proc. Fourth ICCV Workshop Texture Analysis and Synthesis,* pp. 119-124, 2005.

[15] R. Bailey, A. McNamara, N. Sudarsanam, and C. Grimm, "Subtle Gaze Direction," *ACM Trans. Graphics,* vol. 28, no. 4, pp. 1-14, Aug. 2009.

[16] E. Mendez, S. Feiner, and D. Schmalstieg, "Focus and Context in Mixed Reality by Modulating First Order Salient Features," *Proc. Smart Graphics,* 2010.

[17] S. Frintrop, E. Rome, and H.I. Christensen, "Computational Visual Attention Systems and Their Cognitive Foundations," *ACM Trans. Applied Perception,* vol. 7, no. 1, pp. 1-39, Jan. 2010.

[18] L. Itti and C. Koch, "A Model of Saliency-Based Visual Attention for Rapid Scene Analysis," *Analysis and Machine Intelligence,* vol. 20, no. 11, pp. 1254-1259, 1998.

[19] L. Itti, "Models of Bottom-Up Attention and Saliency," *Neurobiology of Attention,* L. Itti, G. Rees, and J. Tsotsos, eds., pp. 576-582, Elsevier, 2005.

[20] J.M. Wolfe, A. Oliva, T.S. Horowitz, S.J. Butcher, and A. Bompas, "Segmentation of Objects from Backgrounds in Visual Search Tasks," *Vision Research,* vol. 42, no. 28, pp. 2985-3004, Dec. 2002.

[21] M.J. Bravo and H. Farid, "Search for a Category Target in Clutter," *Perception,* vol. 33, no. 6, pp. 643-52, Jan. 2004.

[22] J.M. Wolfe and T.S. Horowitz, "What Attributes Guide the Deployment of Visual Attention and How Do They Do It?," *Nature rev. Neuroscience,* vol. 5, no. 6, pp. 495-501, June 2004.

[23] R.A. Rensink, "The Management of Visual Attention in Graphic Displays," *Human Attention in Digital Environments,* C. Roda, ed. Cambridge Univ. Press, 2011.

[24] M. Livingston, J. Gabbard, J.S. II, and C. Sibley, "Basic Perception in Head-Worn Augmented Reality Displays," *Human Factors in Augmented Reality Environments,* W. Huang, L. Alem, and M.A. Livingston, eds., Springer, 2012.

[25] Y. Liu, "Some Phenomena of Seeing Shapes in Design," *Design Studies,* vol. 16, pp. 367-385, 1995.

[26] A. Hoffmann and M. Menozzi, "Applying the Ishihara Test to a PC-Based Screening System," *Displays,* vol. 20, pp. 39-47, 1999.

[27] A.A. Michelson, *Studies in Optics,* p. 208, Dover Publications, 1995.

[28] D. Purves and R.B. Lotto, *Why We See What We Do Redux: A Wholly Empirical Theory of Vision,* second ed., p. 262, Sinauer Assoc., Inc., 2010.