

# A semi-automated approach to balancing of bottom-up saliency for predicting change detection performance

Milan Verma

School of Electronic Engineering and Computer Science,  
Queen Mary, University of London, London, UK



Peter W. McOwan

School of Electronic Engineering and Computer Science,  
Queen Mary, University of London, London, UK



Previous change blindness studies have failed to address the importance of balancing low-level visual saliency when producing experimental stimuli for a change detection task. Therefore, prior results suggesting that top-down processes influence change detection may be contaminated by low-level saliency differences in the stimuli used. Here we present a novel technique for generating semi-automated balanced modifications to a scene, handled by a genetic algorithm coupled with a computational model for bottom-up saliency. The saliency model obtains global saliency values for input images by analysing peaks in feature contrast maps. This quantification approach facilitates the generation of experimental stimuli using natural images and is an extension to a recently investigated approach using only low-level stimuli (Verma & McOwan, 2009). In this exemplar study, subjects were asked to detect changes in a flicker task containing the original scene image (A) and a synthesised modified version (A'). We find under the conditions where global saliency is balanced between A and A' as well as between all modifications (all instantiations of A') that low-level saliency is indeed a reasonable estimator of change detection performance in comparison with high-level measures such as mouse-click densities. When the saliency of the changes are similar, addition/removal changes are detected more readily than colour changes to the scene.

Keywords: change detection, change blindness, visual saliency, visual attention, scene processing

Citation: Verma, M., & McOwan, P. W. (2010). A semi-automated approach to balancing of bottom-up saliency for predicting change detection performance. *Journal of Vision*, 10(6):3, 1–17, <http://www.journalofvision.org/content/10/6/3>, doi:10.1167/10.6.3.

## Introduction

### Introduction to change blindness

When viewing an original image (A) and the modified version (A'), *change blindness* occurs when the viewer cannot detect the change. For example, compare A with A<sub>1</sub>, A<sub>2</sub>, A<sub>3</sub>, or A<sub>4</sub> in [Figure 4](#). Not only can a viewer miss large changes (Grimes, 1996) but they also tend to be entirely confident that they would notice such striking changes (Levin, Drivdahl, Momen, & Beck, 2002). This experimental paradigm has proved to be a powerful tool for exploring various lines of research in visual attention, consciousness, memory, perception, and cognitive science.

The nature and robustness of the effect has led to studies conducted using psychophysics (e.g., Grimes, 1996) as well as real-life interactions (for a summary of techniques, see Rensink, 2002; Simons & Levin, 1997).

In a controlled laboratory setting, a change in the two scenes (A) and (A') can be noticed immediately if the two scenes are spatially aligned and presented in sequence. However, change blindness can occur when viewing the scenes side-by-side or when the scenes are blended together to achieve a gradual change. Another experimental approach was presented by Rensink, O'Regan, and Clark

(1997) known as a *flicker task*, where an interstimulus interval (ISI) is inserted between (A) and (A') as well as after (A'). This sequence is iterated until the observer responds, allowing the accuracy of the response as well as the reaction time to be recorded. Several change detection experiments have been conducted using these approaches to explore the foundations for change blindness. One explanation points to the absence of a sufficient internal representation of the region to be altered (Noë, Pessoa, & Thompson, 2000; O'Regan & Noë, 2001). An alternative explanation is that the pre-change information is encoded but is disrupted, overwritten, or forgotten (Beck & Levin, 2003; Landman, Spekreijse, & Lamme, 2003; Rensink et al., 1997). A third explanation is that the pre-change information is represented and retained but not compared with the post-change representation (Mitroff, Simons, & Levin, 2004; Scott-Brown, Baker, & Orbach, 2000). Since a change region must be localized in order for a change to be detected or sensed (Galpin, Underwood, & Chapman, 2007; Mitroff & Simons, 2002), change blindness can be studied by observing the allocation of visual attention in the scene. The now classic eye-movement evidence by Yarbus (1967) found that saccades and fixations do not follow random paths and are dependent on particular viewing conditions. These viewing patterns have led to the formulation of computational models for visual

attention which can be broadly categorized as being either top-down or bottom-up.

## Top-down versus bottom-up attention

On the one hand, top-down models show that our perception and interpretation of an object is influenced by factors such as prior knowledge (Smith, Hopkins, & Squire, 2006), the context (Torralba, Oliva, Castelhan, & Henderson, 2006), and the task at hand (Triesch, Ballard, Hayhoe, & Sullivan, 2003). According to such theories, the general scene–schema (or *gist*) is obtained after very brief glimpses (Biederman, 1972) and is coupled with knowledge to direct attention around the scene. As an example, when viewing an office environment and searching for a computer, we might expect to find it on the desk rather than on a shelf. We might also expect only certain types of objects that would fit in the general setting of the scene (e.g., a stapler rather than a football on the desk). Object identification and detection has been shown to be more accurate when primed by a semantically consistent scene (Davenport & Potter, 2004; Palmer, 1975). However, objects that violate the scene–schema have been shown to draw earlier and/or longer fixations (Biederman, 1972; Loftus & Mackworth, 1978; Underwood, Templeman, Lamming, & Foulsham, 2008). Unfortunately, the scene-inconsistent object advantage has not been replicated in subsequent scene perception investigations (De Graef, Christiaens, & d’Ydewalle, 1990; Henderson, Weeks, & Hollingworth, 1999; Underwood & Foulsham, 2006; Underwood, Foulsham, van Loon, Humphreys, & Bloyce, 2006; Underwood, Humphreys, & Cross, 2007).

Bottom-up theories, on the other hand, propose that image properties drive the allocation of attention. They can be viewed as a low-level visual search problem where features are analysed in parallel across the visual field and highly contrasting regions that *pop out* are allocated attention. These models are biologically plausible (de Brecht & Saiki, 2006; Itti & Koch, 2000; Koch & Ullman, 1985; Treue, 2003) and can be implemented for a variety of features. Typically, each feature has a feature map, and these feature maps are combined to build a saliency map. The model proposed by Itti and Koch (2000) hypothesized the order of viewer fixations by inhibiting the return of previously fixated areas. This allows the use of the saliency map to traverse from high to low regions drawing attention. These saliency maps match well with early saccadic movements whether participants are encoding pictures in preparation for a memory test (Foulsham & Underwood, 2007; Underwood & Foulsham, 2006; Underwood et al., 2006) or free viewing (Parkhurst, Law, & Niebur, 2002). Parkhurst et al. (2002) have shown that the saliency at fixation is better than a chance fixation distribution. A flickering light or an abrupt onset can

involuntarily capture attention (Posner, 1980), a property widely used by police cars, ambulances, and railroad crossings. This has been found to occur independently of goals and the task at hand (Christ & Abrams, 2006; Mulckhuysen, Van Zoest, & Theeuwes, 2008; Neo & Chua, 2006; Schreij, Owens, & Theeuwes, 2008). Although spatial and temporal feature contrasts could explain this finding, a probabilistic-based formulation has been proposed where salient regions may be defined by the amount of bottom-up Bayesian Surprise (Itti & Baldi, 2006). Contrasting studies have found that this attention capture effect can be modulated by top-down factors (e.g., Lien, Ruthruff, Goodin, & Remington, 2008). Hollingworth and Henderson (1998) suggest that the processing of object information is functionally independent of the scene context. This finding not only conflicts with evidence of a consistent or inconsistent scene advantage but also supports the idea that image feature contrasts objectively draw attention. Recent studies have also found a relationship between image salience and high-level cognitive functions, for tasks involving working memory (Fine & Minnery, 2009) and for scene labeling (Elazary & Itti, 2008). To obtain a more accurate model of scene perception, presumably both bottom-up and top-down factors should be incorporated, and advances have been made to achieve this (Navalpakkam & Itti, 2005).

## Visual attention and change detection

The ongoing debate of top-down versus bottom-up influences also incorporates evidence from change detection studies. Rensink et al. (1997) found that when the transient in the signal accompanied by the change is suppressed by a blank frame, change blindness ensues. Attention is not drawn to the change region, resulting in long change detection times or the change not being detected at all. To explore the influence of high-level effects on attention, Rensink et al. conducted a two-part experiment. The first part involved each participant rating regions of image according to interest; regions selected three or more times were classified as of central interest, and regions not selected at all were deemed to be of marginal interest. The second part tested these classified regions of interest by making changes to them and observing which changes are detected more readily. In order to avoid selection bias, changes were manually controlled for intensity and colour, but the size of the change was on average 4 sq. deg (or 20%) larger for marginal changes. Even though central changes were found to be detected quicker than marginal changes, the subjectivity of the initial interest assessment may have confounded with salience levels. In other words, the “high interest” regions could have been selected because they also had high image salience. This could explain why image feature statistics are highly correlated with semantic

information which is driven endogenously by top-down factors (Henderson, Brockmole, Castelano, & Mack, 2007). However, there have been competing results concerning whether this saliency argument accounts for the finding that changes to areas rated high in subjective interest are more easily detected. Shore and Klein (2000) found that saliency is key in a flicker task. Using the same stimuli as Rensink et al., they rotated half of the image pairs presented by 180° in order to decouple any possible context effects. The original and modified scenes were first displayed side-by-side and participants were asked to identify where a change occurred. When viewing upright images, central changes were detected more readily than marginal changes, in line with the findings of Rensink et al. However, the advantage of central changes was inhibited by inverting the orientation of the scenes, underlining the importance of scene semantics. Shore and Klein found that when using the flicker paradigm, that central and marginal changes are unaffected by scene inversion, instead showing that changes are detected by stimulus-driven processes not scene semantics. Saliency was not measured in the experiments conducted but was instead used for a post hoc explanation of their central versus marginal interest differences. Kelley, Chun, and Chua (2003) on the other hand attempted to manually balance saliency levels by eye, concluding that high-level factors suppressed the influences of low-level saliency. In their experiments, two competing changes were used per scene to address the imbalance of saliency between marginal and central changes. This imbalance could explain why Rensink et al. found a central change advantage and also why Shore and Klein found a null effect of scene inversion. Kelley et al. identified the need to balance the low-level discriminability of changes and therefore manually matched central and marginal changes per scene for size, colour, eccentricity from the centre, and background contrast. However, the study did not control for imbalances in the saliency between the original and modified images, which could have biased attention to the more salient of the two change regions. Manual changes to scenes using photo editing software, especially the addition or removal of objects, could have resulted in the altered region being salient due to differences in luminance or colour balance with its background. Another issue was that Rensink et al., Shore and Klein, and Kelley et al. failed to address the subjectivity of the interest categorization, since participants were always asked to determine the regions of central and marginal interest. In order to objectively assess the image saliency of the changes made, Stirck and Underwood (2007) used an approach which considered intensity, colour, and orientation image feature contrasts. They avoided image artefacts being left by manual changes through taking a photograph of a scene, replacing an object, and taking a second photograph. The Itti and Koch (2000) attention model was used to predict which objects would receive earlier or later fixations. A high saliency change was one which occurred

between one and three in the ordinal ranking and a low saliency change was one which ranked eight or higher. This objectivity also provided a means for fair competition between various object modifications made. They found that change detection performance was positively correlated with scene inconsistency rather than object saliency, in line with the high-level effects observed by Kelley et al. However, the results of Kelley et al. and Stirck and Underwood could have been contaminated since they did not control for the differences in the saliency between the original and modified scenes. The potential imbalance in low-level image properties could have attracted attention to the replaced object, not because of its scene inconsistency but because of its visual saliency. Using a scene pair from the study by Stirck and Underwood and the model described in [Experiment 1](#), [Figure 1](#) shows that a discrepancy exists in saliency map values.

## Balancing saliency levels

To neutralise any effect of differences in low-level image properties between pre-change and post-change scenes as well as for competing changes, our study aims to readdress this problem by using a technique that computationally balances the global saliency levels. To investigate whether the failures to detect changes are due to the change type as well as the saliency of the region being modified, addition/removal and colour changes relating to each scene are examined. In developing our new semi-automated stimulus generation technique, we draw on previous work using line drawings (De Graef et al., 1990; Henderson et al., 1999; Loftus & Mackworth, 1978; Mitroff et al., 2004), computer-generated scenes (Hollingworth & Henderson, 2002), or manual modifications to real-world scenes (Biederman, 1972; Davenport & Potter, 2004; Mitroff & Simons, 2002) to allow controlled modifications. A genetic algorithm is used to drive the process of balancing saliency by searching the alteration space for a modified scene that has a similar global saliency value to the original scene as well as any competing change scenes. Under the viewing conditions of a flicker task, if bottom-up factors guide attention, we would expect the change region as inferred by the saliency map to influence the speed and accuracy to detect changes. Conversely, if top-down factors are an influence, performance should not correlate with changes in high and low saliency regions. Investigating the type of change in these regions could tell us about how representations are encoded and compared.

The first experiment conducted was to gather low-level saliency regions and high-level regions of interest. This was followed by the change blindness experiment. Details of the process of stimulus generation are provided in the corresponding [Stimuli](#) section, followed by the results

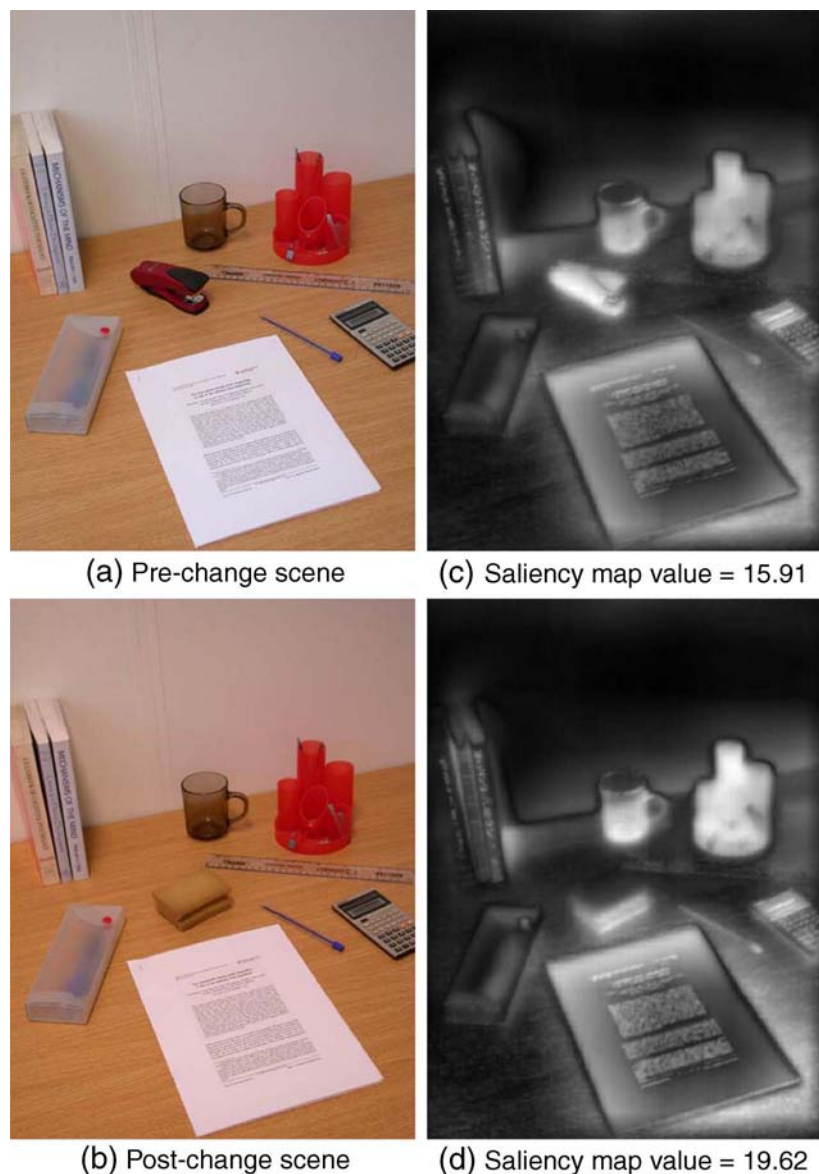


Figure 1. Images from the Stirk and Underwood (2007) study, showing that sometimes an imbalance in salience levels exists, which may introduce confounding variables thus biasing results.

from the flicker experiment using such stimuli. We then suggest possible explanations for the results we have obtained herein.

## Experiment 1: Low-level saliency and interest ratings

### Methods

#### Participants

Sixty-four naïve subjects with ages ranging from 19 to 44 ( $M = 26.71$ ,  $SD = 4.77$ ) participated in the experiments, reporting normal or corrected-to-normal visual acuity.

#### Stimuli

Forty scenes were used, the image dimensions of which were  $640 \times 480$  pixels ( $22.5 \times 16.9$  cm) at a resolution of 72 ppi, subtending visual angles of  $22.3^\circ \times 16.9^\circ$ . The content of the images was a mixture of indoor and outdoor scenes. Some were landscapes images while others were focused on one or two items.

The natural images were colour segmented using mean-shift segmentation (Comaniciu & Meer, 2002) in order to control the content of the scenes for Experiment 2. Any under-segmentation errors violating the semantics of the scene, such as a foreground object containing part of its background, were manually corrected by reassigning the pixels to the correct segment. This segmentation step produces a realistic image, the structure of which is easier to manipulate autonomously.

## Apparatus

The experiment was programmed with the Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997) implemented in MATLAB. Stimuli were presented on a 15-in. monitor at a screen resolution of  $1024 \times 768$ . Participants viewed from approximately 57 cm away from the computer monitor, and movement was restricted using a chin rest. The room was dimly illuminated by a low-intensity light source.

## Procedure

The first stage of the experiment involved obtaining model-predicted low-level saliency measures for each segmented scene, and therefore there was no human interaction for this stage. A low-level saliency map was computed for each scene using the biologically plausible saliency model proposed by Verma and McOwan (2009). In this model, centre-surround similarities and discontinuities for colour, luminance, and orientation

feature contrasts are computed across scales 7 scales according to the size of the input image. There are 2 colour subfeatures (for double-opponent B-Y and R-G channels) and 4 orientation subfeatures ( $0^\circ$ ,  $45^\circ$ ,  $90^\circ$ , and  $135^\circ$ ), and luminance is captured by measuring the brightness of each pixel. The saliency model uses a dynamic feature combination strategy, which logarithmically combines scored subfeature similarities and contrasts according to a peak analysis using Hurst exponent estimations. Once these subfeatures have been fused to form conspicuity maps, these are linearly combined to produce the final saliency map. Since the colour, luminance, and orientation feature maps are normalized and linearly combined, no biasing of features occurs in this methodology. The dynamic weighting procedure ensures that the contribution of each subfeature is never fixed but is instead dependent on the activity peaks of centre-surround feature similarities and contrasts. Stronger isolated peaks are given a high weighting, so these contribute more towards the feature map. Another benefit of the peak analysis process is that the contribution of each subfeature

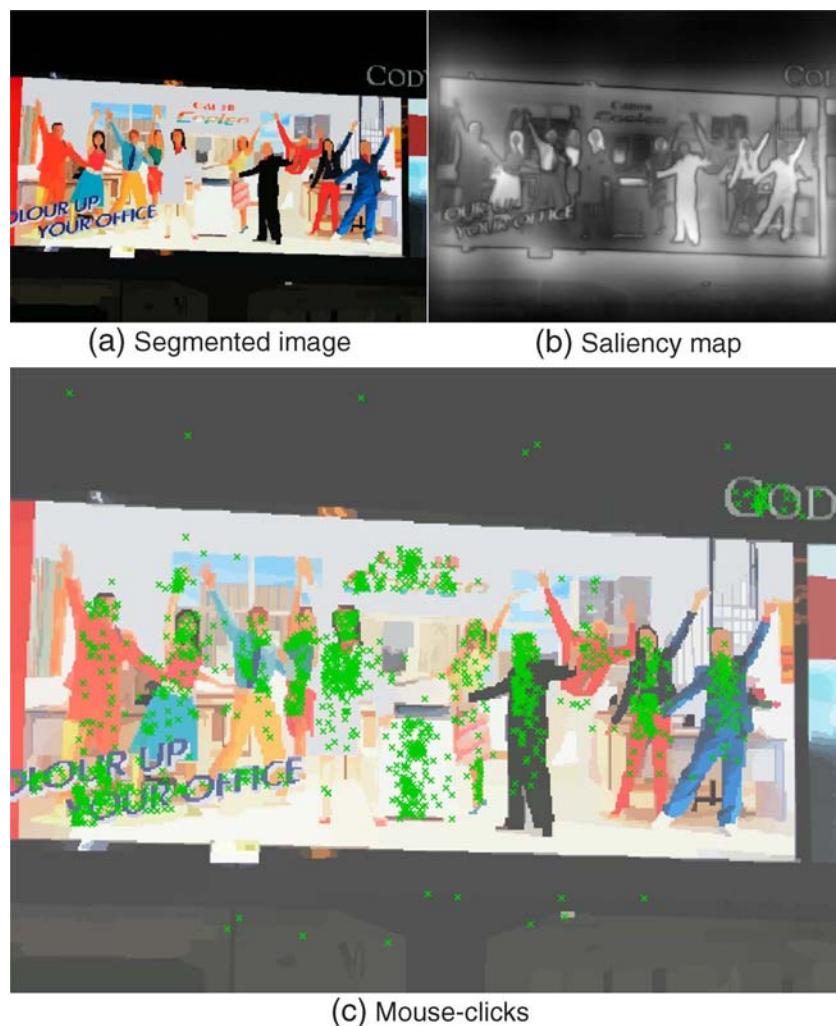


Figure 2. The mouse-click locations in a sample-segmented image selected by 64 subjects are shown in panel c. The observed segmented image (a) and the saliency map (b) are also shown.

conspicuity map can be quantified to form an overall saliency value for each given scene. Specifically, the Hurst exponent is estimated for each centre-surround contrast image computed across spatial scales for colour and luminance subfeatures. The Hurst exponent is also estimated for orientation subfeatures across spatial scales. For each collection of Hurst exponent estimates per subfeature, the maximum is taken to form an accumulative sum. The final sum of these scores is the overall saliency value for the given saliency map, which is particularly important to obtain for [Experiment 2](#). See Verma and McOwan for more details.

Each segmented region acts as a potential focus of attention region. First, a connectivity structure is formed to provide bi-directional linkage information for pixels in the segmentation image to segmented regions in the same image. The pixel intensities of the saliency map are sorted in descending order of saliency. Next, to establish a saliency ordering of segmented regions, each sorted pixel value from the saliency map is matched with its corresponding linked segmented region. Once a high saliency value pixel has been assigned to a region, any subsequent lower value pixels locally present in the saliency map are prohibited from being associated with the same region. This procedure ensures that each segmented region is associated with the highest pixel intensity value for that region. Assuming that these high saliency map pixel values attract attention, this technique crudely replicates the biological behaviour of an inhibition-of-return mechanism. We defined high saliency regions as those among the four highest ranked segments and low saliency regions as those between four and eight in the ranked list.

The second stage of the experiment involved obtaining interest ratings from subjects, and combining these to form a consensus on regions of interest. We defined high and low interest regions according to the segmented regions corresponding to mouse clicks by subjects. All 64 subjects clicked on regions within segmented images that they determined were significant for describing the content of the scene. Even though no time limit was applied, the subjects were only able to click on a maximum of eight pixel locations. The segmented portion(s) relating to densely clicked areas showing inter-subject agreement for each scene was defined as highly interesting. As with the classification of salient regions, high interest regions were the top four regions and low interest regions were ranked fifth to eighth.

## Results

Segmented interest regions from mouse-click densities and segmented salient regions were compared by conducting an ANOVA on the corresponding ordinal rank data. However, no statistically significant association was found between mouse-click densities and salient regions

at the  $p < 0.05$  level. This shows a disparity between model-predicted low-level saliency regions and interest regions. [Figure 2](#) presents a sample-segmented image presented to participants and the corresponding interest points for all 64 subjects. The figure illustrates the difference between the saliency model inferred locations and the areas of high-density mouse clicks. In particular, notice that the blue text and the white printer are not highlighted by the saliency model yet they are identified as interest points through mouse clicks.

## Experiment 2: Change detection

### Methods

#### Participants

Sixty-four naïve subjects with ages ranging from 19 to 37 ( $M = 25.74$ ,  $SD = 4.23$ ) participated in the experiments, none of whom participated in [Experiment 1](#). All subjects reported normal or corrected-to-normal visual acuity.

#### Stimuli

The same forty colour-segmented scenes were used from [Experiment 1](#). The image dimensions of which were  $640 \times 480$  pixels ( $22.5 \times 16.9$  cm) at a resolution of 72 ppi, subtending visual angles of  $22.3^\circ \times 16.9^\circ$ .

The aim of this experiment was to test the hypothesis that high change detection rates are linked to low-level salient regions rather than regions of interest. The novelty of our approach is in the use of an automated system for modifying scenes and balancing saliency levels. These modified scenes will then be used for observing change detection performance, monitoring change detection rates with respect to the location of the change and the type of the change.

Visual attention may be directed by bottom-up (exogenous, stimulus-based) or top-down (endogenous, goal-directed) control. Thus, to decouple the influence of salience and visual interest through scene semantics, the 40 scenes were modified in high interest and low interest regions and well as high and low saliency regions, which were determined in [Experiment 1](#). Two changes were made per scene without human intervention, for each of the four location types. These two changes were colour and addition/removal changes.

The processing pipeline for automated scene modification is shown in [Figure 3](#). The procedure was facilitated by a genetic algorithm (GA) (Davis, 1991), where the saliency model was used as the fitness function. For each scene, a variable-length chromosome held HSV values for all segmented regions, including a collection of bits denoting the candidate change region for that chromosome. In addition, the chromosome also held the saliency

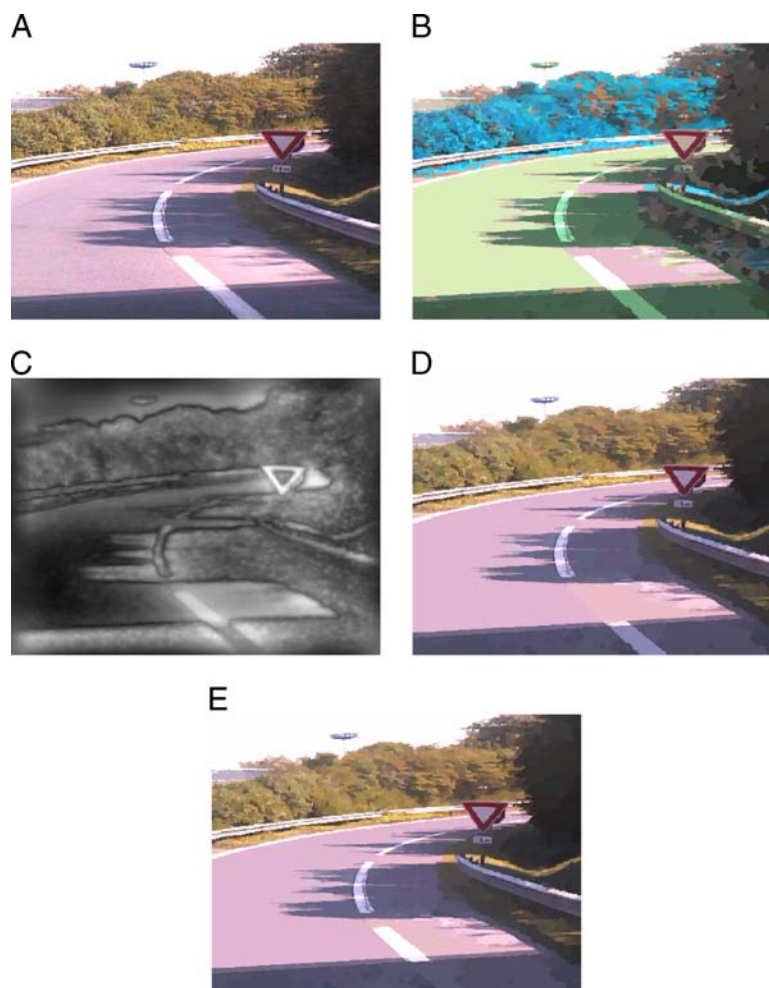


Figure 3. The processing pipeline for stimulus generation. The original image (a) is processed to produce the mean-shift segmented image (b) which is shown here with colour-adjusted regions to illustrate that contiguous homogeneous pixels are grouped together. This image is then used to produce a grayscale saliency map (c). Using this, a GA is then used to suggest changes in high (d) and low (e) saliency regions. In this particular case, the high saliency change is the removed road marking and the low saliency change is the removed roadside safety barrier. The same approach is used for modifying interest regions as determined in [Experiment 1](#).

value of its counter-part pre-change scene. The fitness function was responsible for comparing the changed scene saliency level with the stored pre-change saliency level and ensuring that only scene pairs with the smallest two saliency differences in the population ( $N = 12$ ) were selected. Accepted scenes were forwarded to the next generation with the ultimate aim of the evolutionary process being to minimize this saliency error for the changed region. The termination criterion was a tolerance level which ensured that computed global saliency values for all 40 scenes were within  $\pm 3.0$  units ( $M = 34.4$ ,  $SD = 7.8$ ). A narrower margin may be used; however, this may decrease the convergence rates and limit the range of the resultant candidate changes.

The role of the GA was to optimize the search for changes by minimizing the computed global saliency values between each original and modified scene. [Figure 4](#) shows the schema for the balancing checks made by the

GA. To prevent an imbalance between the competing changes due to this tolerance and also due to the variability within the high and low saliency groups used, the saliency value of colour and addition/removal change scenes were also balanced. The purpose of balancing saliency in this rigorous manner was to avoid any attentional bias that could be caused by an imbalance in image saliency. Elitism was set to 0.2, one-point crossover and mutation was used, the crossover rate was 0.80, the mutation rate was 0.08, and the extent of each HSV mutation was limited to 5%.

Colour changes were produced by manipulating HSV value bits, ranging from  $0^\circ$  to  $360^\circ$  for H and from 0.0 to 1.0 for the S and V components. Removal changes were produced by matching the change region colour with the neighbouring background colour. In terms of the GA, the type of change to seek was regulated by an immutable change type bit—0 denotes colour change, 1 denotes object

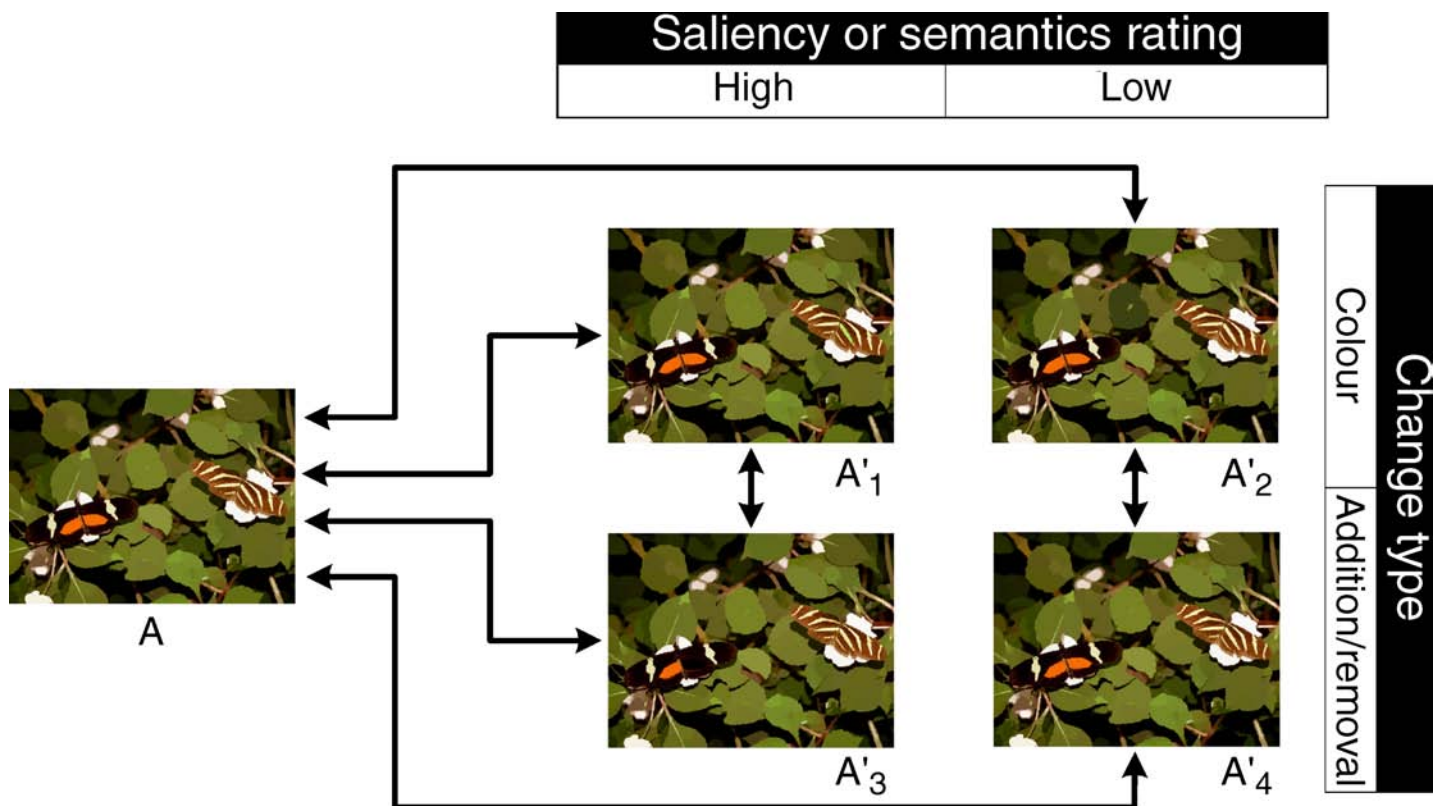


Figure 4. The schema for balancing the global scene saliency values. Saliency was balanced across original and modified scenes ( $A$  vs.  $\{A_1, A_2, A_3, A_4\}$ ). The genetic algorithm may obtain an optimal but not always exactly matching modification. For fair competition, competing changes ( $A_1$  vs.  $A_3$  and  $A_2$  vs.  $A_4$ ) were also balanced. An example of an original scene (i.e.,  $A$ ) is image a in Figure 3 and examples of modified scenes such as  $A_3$  and  $A_4$  are images d and e in Figure 3, respectively.

addition, and 2 denotes object removal. The background of a given removed item was reconstructed by using a dithering approach, which sampled four nearby pixels and used the mean colour values to fill the space. The four pixels were randomly chosen along the  $x$  and  $y$  planes of the segmented region, and these were encoded into the candidate change chromosome. The removal of a segmented region requires the largest number of generations to optimize due to the mismatch in saliency levels. Using this approach, it is plausible that some object removals could be drastic enough to never converge on a satisfactory changed-scene counterpart. Such images were not used in our experiments. Given a pair of scenes in which an object has been removed, the effect of object addition was created by switching the original and modified scenes, for instance, displaying Figure 3d before the non-colour-adjusted version of Figure 3b. Figure 5 presents examples of candidate modifications in a low-saliency region, along with their corresponding saliency maps and numeral saliency values. Figure 5b shows how the scene appears when the region is removed, Figure 5c shows how the scene appears when the region colour is altered, and Figure 5d shows an example of a region change which is above the acceptable tolerance level.

### Apparatus

The same equipment and setup was used as those in Experiment 1. Subjects responded to stimuli by pressing assigned keys on a modified keyboard.

### Procedure

Modifying the 40 original scenes produced 320 different possible original-changed pairs, 160 images per level of the change-type independent variable (colour and addition/removal). This included 40 scenes each for the two levels of saliency and two levels of interest (high and low). To avoid the same scene being presented twice to a participant, eight groups of 40 unique original-changed pairs were formed, which were run on four separate subject groups. Each design condition (e.g., high interest colour change) was shuffled across the eight groups in order to test a variety of conditions per group. There were eight participants per group, so each version of an original scene was shown eight times. Participants were given written instructions to prepare for a memory task by observing a set of paired images separated by an ISI containing a blank grey frame. They were informed that changes may occur between the pairs of images presented



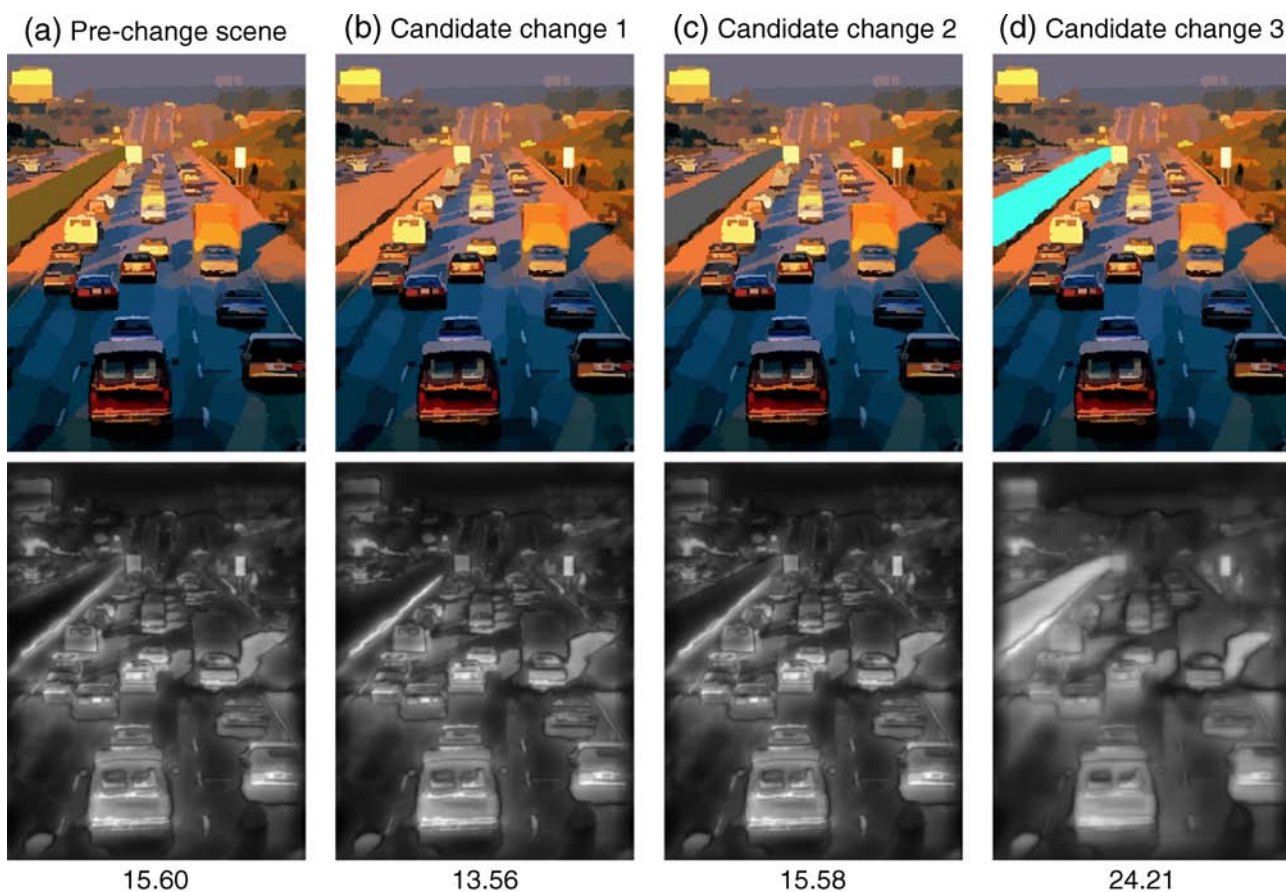


Figure 5. Examples of scene modifications suggested by our automated system. The first row shows the pre-change scene and three types of changes. The second row shows the corresponding saliency maps and their numeral value. Notice that the difference in the saliency values of candidate change 3 and the pre-change scene is more than the acceptable level of 3.0 units. For this reason, the candidate change is rejected by the fitness function, whereas candidate changes 1 and 2 are accepted.

and that there would only be one object change per pair of images. Participants were asked to decide whether a change between the scenes had occurred. A practice block

of six trials was run before the experiment began. The trial structure is presented in Figure 6. Each scene was repeatedly presented until the participant responded with

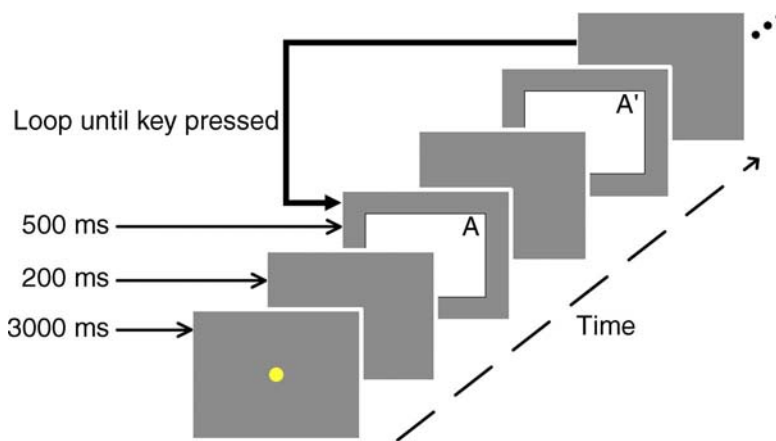


Figure 6. The trial structure for Experiment 2. Participants were presented with 80 scenes in which, either no change was made, or changes were made to the colour or the presence of an object. The region in which the change was made was determined either by a saliency map or by a mean subject rating gathered prior to the experiment. Participants were asked to decide as to whether or not a change occurred.



Movie 1. A Quicktime example of a synthesised modification using the procedures described in this paper.

a key press indicating “change” or “no change.” The assignment of key location was randomized between subjects, and the keys were modified to mask their labeling. To initiate a trial, participants fixated on a yellow circle presented in the middle a grey frame for 3000 ms, followed by an ISI containing a blank grey frame (without the fixation circle), which was presented for 200 ms. This was followed by the first image of a scene pairing against a grey background for 500 ms. Next, a blank ISI was presented again for 200 ms before the second image of the pairing was presented for 500 ms. This second image was either the same as the first image or a modified version. To complete the cycle, there was another ISI for 200 ms before the sequence was repeated. The timer was started upon the onset of the first image, and the response time (RT) was measured from the onset of the second image. Any response made before the first 700 ms in the presentation cycle was ignored. Once a response was made, the sequence was stopped and the next trial was initiated. For each trial, both the change region and the change type were set according to the eight subject groups detailed above. Both the scene order and whether the image pair was original-change or change-original was randomized. Finally, half the trials presented

contained no change, resulting in a total of 80 trials per participant. [Movie 1](#) presents an example of a stimulus image following the mentioned trial structure.

## Results

A breakdown of the mean response times is presented in [Table 1](#). The mean response time for changes in regions of high saliency was 3207 ms, compared with 8132.5 ms for low saliency regions. The mean response time for changes in regions of high interest was 5014.5 ms compared with 6898.5 ms for low interest regions. Performance results show that overall, changes in high saliency regions tend to be detected quicker when an object is added or removed. This effect is present, but reduced, in regions of low saliency, where colour changes are detected only marginally slower. Modifications to high regions of interest are detected quicker than low regions of interest for both types of changes. Also, for high interest region changes, the response times are generally higher for both types of changes as compared with highly salient region modifications. For low interest regions, both types of changes are detected slightly faster as compared with low salient region changes.

These results indicate that changes made to high saliency regions are detected faster than changes to low saliency regions. Changes to high saliency regions are also detected much faster than regions rated to be of high interest, suggesting that perhaps saliency has a greater pre-attentive contribution to building an internal representation of the scene and for change detection. Overall, addition/removal changes were detected quicker than colour changes, regardless of the area in which the change was made. In particular, addition/removal changes in high saliency regions were detected far quicker than colour changes. The mean responses for addition/removal and colour changes in low saliency regions or regions of low interest were similar, indicating that the type of change had no significance in these regions.

Mean response times were subjected to a three-way within-subjects analysis of variance (ANOVA): Region Selection (saliency, interest)  $\times$  Region Type (high, low)  $\times$  Change Type (colour, add/remove). Trials in which no change occurred were excluded from the analysis along

| Region    |      | Colour        | Addition/removal | Averages |
|-----------|------|---------------|------------------|----------|
| Saliency  | High | 3681 (900.2)  | 2733 (430.3)     | 3207     |
|           | Low  | 8276 (3967.2) | 7989 (3313.1)    | 8132.5   |
| Semantics | High | 5696 (632.3)  | 4333 (680.7)     | 5014.5   |
|           | Low  | 7223 (965.2)  | 6574 (1345.7)    | 6898.5   |

Table 1. Mean response times (in ms) for detecting changes in high/low saliency or interest (semantics) regions as a function change type (SDs in parentheses).

with RT outliers more than two *SDs* away from the subject's mean. Analysis of reaction time data showed that there was a main effect of Region Selection,  $F(1,63) = 26.56$ ,  $p < 0.001$ , showing that changes made to salient regions were detected quicker than changes made to regions of interest. There was also a significant main effect of Region Type,  $F(1,63) = 70.31$ ,  $p < 0.001$ , which indicated that, overall, changes to regions of high interest/saliency were detected quicker than changes to regions of low interest/saliency. A small but significant main effect of change type was also observed,  $F(1, 63) = 5.34$ ,  $p < 0.05$ , with colour changes taking longer to detect than addition and removal of objects. A significant Region

Type  $\times$  Change Type interaction effect was also observed,  $F(1,63) = 5.64$ ,  $p < 0.05$ , as well as an interaction between Region Type  $\times$  Region Selection,  $F(1,63) = 13.64$ ,  $p < 0.05$ . In particular, the regions selected using the saliency model showed a robust effect of Region Type,  $F(1,63) = 27.46$ ,  $p < 0.005$ , whereas subjectively chosen regions showed no effect of Region Type,  $F(1,63) = 0.24$ , *ns*. In summary, these results indicate that changes in subjectively rated regions of interest take longer to detect than objectively selected regions using a saliency model.

Analysis of the accuracy data also revealed significant main effects of Change Type,  $F(1,31) = 15.55$ ,  $p < 0.01$ , and Region Type,  $F(1,31) = 28.33$ ,  $p < 0.01$ . As with RT data, the latter main effect was modulated by two reliably significant interactions of Change Type,  $F(1,31) = 22.32$ ,  $p < 0.01$ , and Region Selection,  $F(1,31) = 35.34$ ,  $p < 0.01$ . Mean accuracy percentages (Figure 7) show that participants made more correct responses in high saliency regions than low ones (86.65% vs. 74.95%) and in high interest regions than low ones (74.7% vs. 71.25%). In high saliency regions, responses were also more accurate for addition/removal changes than colour changes (88.9% vs. 84.4%). Similarly, addition/removal changes in high interest regions were detected with greater accuracy than colour changes (76.9% vs. 72.5%). Addition/removal change accuracy was marginally better than for colour changes in low saliency regions (76.6% vs. 73.3%). A similar result was observed for addition/removal versus colour change accuracy (72.6% vs. 69.9%). These results were significant at the 0.01 level. There was no evidence of a speed-accuracy trade-off.

A third experiment was conducted to assess eye-tracking data for pre-change and post-change scenes. Since global saliency is balanced for these scene pairs, fixation durations at modified regions should be similar.

## Experiment 3: Saliency levels versus eye-tracking data

### Methods

#### Participants

Thirteen naïve subjects with ages ranging from 22 to 37 ( $M = 28$ ,  $SD = 4.02$ ) participated in the experiments, reporting normal or corrected-to-normal visual acuity.

#### Stimuli

Each participant was presented with thirty-two images in this experiment; half were pre-change scenes and the other half were post-change scenes. The post-change scenes contained two images per change type (high saliency colour change, high saliency addition/removal change, low

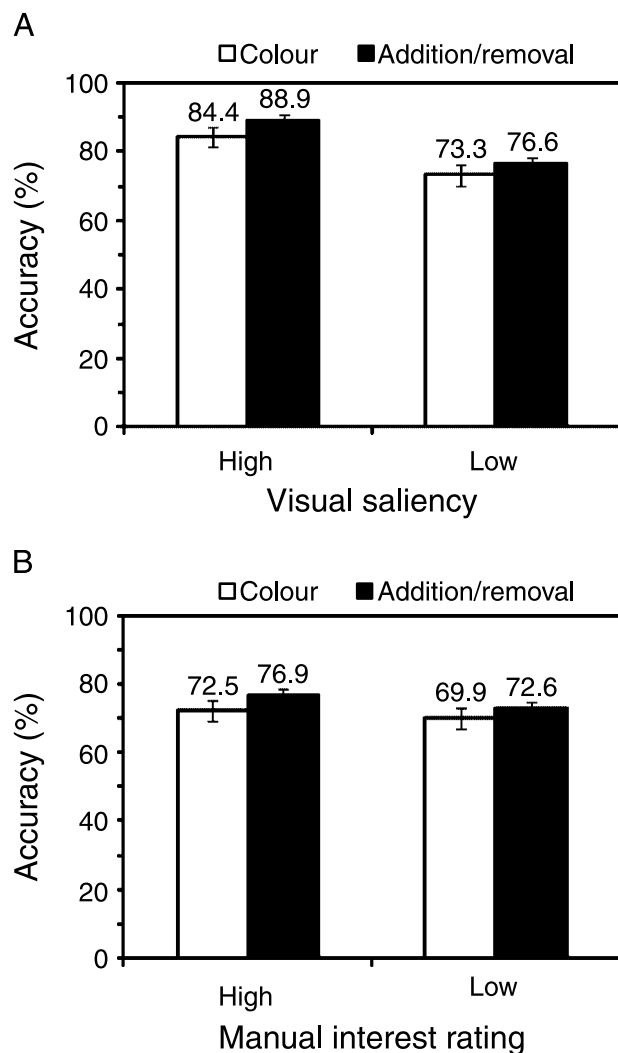


Figure 7. Accuracy data for detecting colour and addition/removal-based changes to model-predicted regions of interest (top) and manually rated interest regions (bottom). Observers readily detected changes made to regions of high saliency or high interest, particularly when the change influenced the existence of an object in the scene. Overall, accuracy was higher for objective model-determined regions as opposed to subjectively selected regions. Error bars indicate 1 *SEM*.

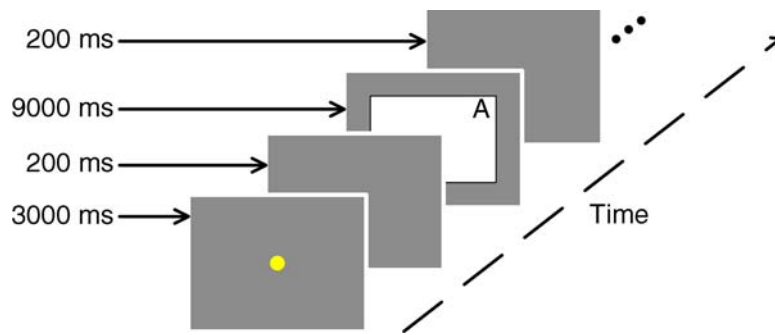


Figure 8. The trial structure for [Experiment 3](#) and [Experiment 4](#). Image A is either a pre-change or a post-change scene.

saliency colour change, and low saliency addition/removal change). As with [Experiment 1](#), we defined high saliency regions as those among the four highest ranked segments and low saliency regions as those between four and eight in the ranked list. The image dimensions were also the same as the preceding experiments,  $640 \times 480$  pixels ( $22.5 \times 16.9$  cm) at a resolution of 72 ppi. These images subtended visual angles of  $22.3^\circ \times 16.9^\circ$ . The images presented were of unique scenes in order to avoid any high-level influences.

### Apparatus

The CRS 50 Hz video eye-tracker coupled with the MATLAB Video Eyetracker Toolbox were used to record eye movements. Stimuli were presented on a 15-in. monitor at a screen resolution of  $1024 \times 768$ . Participants viewed from approximately 57 cm away from the computer monitor, and movement was restricted using a chin rest. The room was dimly illuminated by a low-intensity light source. Eye-tracking data were gathered for both pre-change and post-change scenes at altered regions.

### Procedure

The experiment began with a nine-point calibration step. Participants were asked to freely view images, which were a mixture of pre- and post-change scenes. The change region coordinates were recorded so that fixation data within these regions could be compared between pre-

and post-change scenes. The trial structure is shown in [Figure 8](#).

## Results

For this experiment, only the fixation data at the change region were analysed. Fixations less than 50 ms were removed from the analysis to avoid using unreliable data from brief fixations. Mean fixations numbers and durations are presented in [Table 2](#). The number of fixations between the onset of the stimulus image and first fixation of either a high or a low saliency change region was recorded. This was used to indicate the conspicuity of the region to draw the viewers' attention. The fixation duration was taken as an indication of how complex the region was to interpret.

Mean fixations were subjected to a two-way within-subjects ANOVA: Region Type (high, low)  $\times$  Scene Type (pre-change, post-change). A main effect of Region Type was observed,  $F(1,12) = 12.97, p < 0.01$ , where regions of high salience were attended earlier than regions of low salience. No statistically significant main effect was observed for scene type, and no interaction effects were observed between scene type and region type ( $F < 1$ ).

Mean fixation durations for the first fixation on a given region were also subjected to a similar ANOVA. However, no main effects were found for Region Type,  $F(1,12) = 4.74, ns$ , or Scene Type,  $F(1,12) = 3.53, ns$ , or an interaction between the two,  $F(1,12) = 5.23, ns$ .

|                                     | Saliency | Pre-change      | Post-change     |
|-------------------------------------|----------|-----------------|-----------------|
| No. fixations prior to 1st fixation | High     | 3.21 (3.32)     | 2.74 (4.16)     |
|                                     | Low      | 8.30 (3.41)     | 8.98 (2.42)     |
| Duration of 1st fixation (ms)       | High     | 257.76 (146.76) | 268.12 (158.30) |
|                                     | Low      | 368.41 (255.45) | 486.01 (165.70) |

Table 2. Mean number of fixations and durations from [Experiment 3](#), as a function of region salience in pre-change and post-change scenes (*SDs* in parentheses).

This experiment empirically validates the usage of the saliency model as a method for predicting early visual attention behaviour, particularly for the preceding change blindness experiments. This is because the experiment shows that model predicted salience corresponds with eye movement data. However, the question arises of whether altering a scene can significantly alter viewing behaviour. Therefore, a fourth experiment was conducted to measure the impact of altering the content at attended regions. This would statistically confirm whether visual attention is controllable using a modified version of our proposed framework. The saliency levels of candidate changes were not balanced in the following experiment.

## Experiment 4: Manipulating saliency levels without balancing

### Methods

#### Participants

Thirteen naïve subjects with ages ranging from 23 to 39 ( $M = 28.62$ ,  $SD = 4.15$ ) participated in the experiments, reporting normal or corrected-to-normal visual acuity.

#### Stimuli

Each participant was presented with twenty-four images in this experiment; half were pre-change scenes and the other half were post-change scenes. Half of the changes made were colour changes and the other half were addition/removal changes. These images contained regions of high salience that had been adjusted to obtain lower global saliency levels. In other words, high saliency regions were transformed into low saliency regions. The image dimensions were the same as the preceding experiments,  $640 \times 480$  pixels ( $22.5 \times 16.9$  cm) at a resolution of 72 ppi. These images subtended visual angles of  $22.3^\circ \times 16.9^\circ$ . The images presented were of unique scenes in order to avoid any high-level influences.

#### Apparatus

The same equipment and setup was used as those in [Experiment 3](#).

### Procedure

The experiment began with a nine-point calibration step. Participants were asked to freely view images, which were a mixture of pre- and post-change scenes. The trial structure for this experiment is shown in [Figure 8](#).

### Results

As with the previous experiment, only the fixation data at the change region were analysed. Fixations less than 50 ms were removed from the analysis to avoid using unreliable data from brief fixations. Mean fixations numbers and durations are presented in [Table 3](#).

Mean fixations were subjected to a two-way within-subjects ANOVA: Region Type (high, low)  $\times$  Scene Type (pre-change, post-change). There was a main effect of Region Type,  $F(1,12) = 9.42$ ,  $p < 0.01$ , with fewer fixations prior to a high saliency region compared with a low saliency region for pre-change scenes. There was also a main effect of region type for post-change scenes, with fewer fixations prior to the first fixation in the low saliency region as compared with the high saliency region,  $F(1,12) = 8.27$ ,  $p < 0.01$ . The number of fixations prior to fixating a high/low saliency region was similar for both pre-change and post-change scenes,  $F(1,12) = 0.44$ ,  $ns$ .

Mean fixation durations for the first fixation on a given region were also subjected to a similar ANOVA. However, no main effects were found for Region Type,  $F(1,12) = 2.22$ ,  $ns$ , or Scene Type,  $F(1,12) = 1.24$ ,  $ns$ , or an interaction between the two factors,  $F(1,12) = 2.70$ ,  $ns$ .

Analysis of the fixation data shows that the time taken to attend highly salient regions is shorter than for regions of low salience. However, when the post-change scenes have the high and low salient regions switched, this pattern is also reversed. [Experiment 3](#) showed that when such changes are made while balancing the saliency levels for pre- and post-change scenes, visual attention behaviour does not differ, this result being confirmed by the eye-tracking data. In contrast, this experiment shows that the visual attention behaviour can indeed be manipulated when there is an imbalance in saliency levels. The number of fixations prior to fixating a region tends to depend on the visual properties of that region. When the region is of low salience, the number of fixations is large. As the salience level for this region is increased, the number of fixations (until this region is attended) is seen

|                                     | Saliency | Pre-change      | Post-change     |
|-------------------------------------|----------|-----------------|-----------------|
| No. fixations prior to 1st fixation | High     | 2.31 (1.39)     | 6.74 (3.34)     |
|                                     | Low      | 7.56 (3.21)     | 3.45 (2.10)     |
| Duration of 1st fixation (ms)       | High     | 322.45 (113.64) | 238.82 (331.35) |
|                                     | Low      | 220.41 (214.24) | 342.97 (180.73) |

Table 3. Mean number of fixations and durations from [Experiment 4](#), as a function of region salience in pre-change and post-change scenes ( $SDs$  in parentheses).

to significantly drop. This is as expected if the model's saliency levels are a valid indicator of early visual attention.

## Discussion

Recently, visual attention theories supporting bottom-up influences have been challenged by new evidence for top-down processes. top-down factors can influence scene perception; however, their influence was thought to occur later in scene viewing as indicated by a decrease in saliency over multiple fixations (Parkhurst et al., 2002; however, for a counter argument, see Tatler, Baddeley, and Gilchrist, 2005). Evidence of these top-down factors modulating eye movements in early vision has forced researchers to re-evaluate the role of bottom-up processes. Here, our studies use the change blindness paradigm to reveal findings supporting the cognitive construction and use of a saliency map. We show that even when top-down information is available that low-level saliency plays an important role in change detection performance.

Due to the brief exposure times prominent in a flicker task, an observer is more reliant on bottom-up visual mechanisms (Shore & Klein, 2000) than in a simultaneously display task. These mechanisms, however, could be misguided by imbalances in visual saliency between the original and modified scenes, providing unreliable change detection results. The drawback of manually and subjectively scoring saliency was addressed by Stirk and Underwood (2007), but there are two issues with their approach that could provide explanations for the reported null effect of saliency. Firstly, the ordinal ranking does not guarantee a balance of saliency between the competing changes. For a given scene, even though both an inconsistent and a consistent object can be ranked as 1st, one change (e.g., the inconsistent change) could be more salient than the other. The more salient object will therefore attract earlier fixations. This issue is exaggerated by the use of high and low saliency groups for balancing changes. A second issue is the saliency of the original scene—what was the saliency level of the replaced object? This is important because two highly salient replacements could have different change detection rates because one replacement is more similar to the original than the other. Stirk and Underwood only evaluated the saliency of the modified scene and balanced saliency across the competing changes. The first issue highlights an imbalance between *competing changes* and the second issue highlights an imbalance between the *original and the modified scene*. This study extends their work by balancing saliency for both of these cases. Doing so reduces the chance that visual transients will attract attention to modified objects and avoids leading to unreliable conclusions for change detection.

Our goal was to assess whether top-down factors triumph bottom-up saliency, when top-down information

is available. Saliency maps were computed and balanced using a solitary global quantifiable measure from our biologically plausible model (Verma & McOwan, 2009). When balanced in this way, our findings show that detection performance can be inferred using low-level models of saliency. Changes made to regions of high saliency were detected more readily than changes to regions of low saliency. Furthermore, colour changes in high saliency regions were more difficult to detect than addition/removal changes, requiring on average 948 more milliseconds to see the change. Indeed, for some images, participants required more than 15 seconds to see the change. Even though the corresponding colour changes in low saliency regions on average took longer to detect (4.6 more seconds), the change type was not as influential in these regions. It follows that the difficulty to detect changes in these regions is not due to low-level image property similarities, nor for that matter, scene consistency. The difficulty does however depend on the region in which the change is made and to a lesser extent the type of change made. Our findings support two possible explanations for why change blindness occurs. Firstly, that allocation of attention is seldom allocated to these low saliency regions for long enough durations to capture and consolidate an object representation. This explanation suggests that encoding failures account for change blindness. Secondly, even when an object representation is maintained, the relevant structures involved for the comparison arguably process change types in a different manner. In particular, that colour changes are processed separately of addition/removal changes. This explanation is supported by physiological studies, finding that colour receptive fields are larger than others therefore requiring larger areas to be activated (Livingstone & Hubel, 1984). This is also consistent with the spatial resolution of colour being relatively low (Livingstone & Hubel, 1987; Mullen, 1985). The colour changes observed in this study manipulate the object while it still exists in the scene, whereas adding or removing an object from the scene could alter any cognitive representation built of the scene. The functional independence of content (colour) and structural (addition/removal) processing and the advantage of structure-based changes observed here suggest that a structural representation of the scene could be captured in early vision. However, this does not rule out the gist of the scene being inferred early on, perhaps using the configuration of the objects in the scene to achieve this. If the structure of the scene is captured early on in viewing, this could explain why structural changes are detected faster than other types of changes. Individually investigating structural and content changes may provide a greater insight into change detection performance, and this could be the focus for future research.

Our findings show that the low-level image saliency map provides an accurate estimation of visual attention in a change detection task, which contrast the findings of Wright (2005). Wright tested subjective and objective

measures of salience and found that subjective measures of salience, taken from subjects pre-selecting salient regions, were more predictive of change detection. Several different spatiotemporal contrast changes were measured as a single objective measure of saliency and were found to not show a relationship with change detection performance, crucially finding that they were also not a good predictor of salience of the changed object. The low-level features we use in this paper appear to provide a change detection prediction performance advantage over measures chosen by Wright. Our findings show that low-level saliency measures are not only a good predictor of visual attention for low-level images (Verma & McOwan, 2009) but also for change detection (Wright, Green, & Baker, 2001). Foulsham and Underwood (2007) have also found that the saliency map predicts fixation locations in a memory task but found this to not be the case when participants are performing a search task. The memory task has a specific role; it encourages viewing behaviour akin to many real-world interaction scenarios. Objects are neutrally analysed unlike a search task that preferentially provides attentional bias to particular targets. Finally, additional support for our findings is provided by a more recent study by Foulsham and Underwood where they found that salient regions were strongly correlated with fixation locations. This correlation was better than a biased or chance fixation model even though the order of fixations was not predictive of human scan paths.

Alternative saliency models may provide conflicting results to ours, perhaps due to their definition of salience. However, the saliency balancing procedure will lead to more reliable data with which more accurate conclusions can be drawn. The novel stimuli generation methodology we have presented in this paper allows rapid semi-automated production of a corpus of synthetic stimuli pairs with prescribed saliency distributions. The experimental results presented show that these stimuli provide useful and convenient customized psychophysical probes and that this technique could prove a useful tool for further study of the change blindness phenomena.

## Conclusions

Although flicker delocalizes motion signals resulting from a change, manual changes to a scene can introduce imbalances in low-level image properties. This paper presents a novel approach for generating change detection stimuli by computationally balancing saliency levels not just for the changed object but for the entire scene. Results show that faster response times and higher detection accuracy are linked with bottom-up saliency and to a lesser extent the change type. Changes made to add or remove an object from the scene are detected more readily than changes to the colour of an object, which retains its existence in the scene.

## Acknowledgments

We thank all the participants in this study for generously assisting us with this research. This work is funded by EPSRC. We would also like to thank Jeremy Wolfe, Mark Becker, Michael Proulx, and Ron Rensink for their insightful comments on the early drafts of the paper. Many thanks also to Dan Simons for helpful discussions.

Commercial relationships: none.

Corresponding author: Milan Verma.

Email: milan@dcs.qmul.ac.uk.

Address: Mile End Road, London E1 4NS, UK.

## References

- Beck, M. R., & Levin, D. T. (2003). The role of representational volatility in recognizing pre- and postchange objects. *Perception & Psychophysics*, *65*, 458–468. [PubMed] [Article]
- Biederman, I. (1972). Perceiving real-world scenes. *Science*, *177*, 77–80. [PubMed]
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*, 433–436. [PubMed]
- Christ, S. E., & Abrams, R. A. (2006). Abrupt onsets cannot be ignored. *Psychonomic Bulletin & Review*, *13*, 875–880. [PubMed] [Article]
- Comaniciu, D., & Meer, P. (2002). Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *24*, 603–619.
- Davenport, J. L., & Potter, M. C. (2004). Scene consistency in object and background perception. *Psychological Science*, *15*, 559–564. [PubMed]
- Davis, L. (1991). *Handbook of genetic algorithms*. New York: Van Nostrand Reinhold.
- de Brecht, M., & Saiki, J. (2006). A neural network implementation of a saliency map model. *Neural Networks*, *19*, 1467–1474. [PubMed]
- De Graef, P., Christiaens, D., & d'Ydewalle, G. (1990). Perceptual effect of scene context on object identification. *Psychological Research*, *52*, 317–329. [PubMed]
- Elazary, L., & Itti, L. (2008). Interesting objects are visually salient. *Journal of Vision*, *8*(3):3, 1–15, <http://www.journalofvision.org/content/8/3/3>, doi:10.1167/8.3.3. [PubMed] [Article]
- Fine, M. S., & Minnery, B. S. (2009). Visual salience affects performance in a working memory task. *Journal of Neuroscience*, *29*, 8016–8021. [PubMed] [Article]

- Foulsham, T., & Underwood, G. (2007). How does the purpose of inspection influence the potency of visual saliency in scene perception? *Perception, 36*, 1123–1138. [PubMed]
- Galpin, A., Underwood, G., & Chapman, P. (2007). Sensing without seeing in comparative visual search. *Consciousness and Cognition, 17*, 672–687. [PubMed]
- Grimes, J. (1996). On the failure to detect changes in scenes across saccades. In K. Akins (Ed.), *Perception: Vancouver studies in cognitive science* (vol. 5, pp. 89–110). New York: Oxford University Press.
- Henderson, J. M., Brockmole, J. R., Castelano, M. S., & Mack, M. (2007). Visual saliency does not account for eye movements during visual search in real-world scenes. In R. van Gompel, M. Fischer, W. Murray, & R. Hill (Eds.), *Eye movements: A window on mind and brain* (pp. 537–562). Oxford: Elsevier.
- Henderson, J. M., Weeks, P. A., & 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.
- Hollingworth, A., & Henderson, J. M. (1998). Does consistent scene context facilitate object detection. *Journal of Experimental Psychology: General, 127*, 398–415. [PubMed]
- Hollingworth, A., & Henderson, J. M. (2002). Accurate visual memory for previously attended objects in natural scenes. *Journal of Experimental Psychology: Human Perception and Performance, 28*, 113–136.
- Itti, L., & Baldi, P. (2006). Bayesian surprise attracts human attention. *Advances in Neural Information Processing Systems (NIPS 2005), 19*, 1–8.
- Itti, L., & Koch, C. (2000). A saliency-based mechanism for overt and covert shifts of visual attention. *Vision Research, 40*, 1489–1506. [PubMed]
- Kelley, T. A., Chun, M. M., & Chua, K. P. (2003). Effects of scene inversion on change detection of targets matched for visual salience. *Journal of Vision, 3*(1):1, 1–5, <http://www.journalofvision.org/content/3/1/1>, doi:10.1167/3.1.1. [PubMed] [Article]
- Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: Towards the underlying neural circuitry. *Human Neurobiology, 4*, 219–227. [PubMed]
- Landman, R., Spekreijse, H., & Lamme, V. A. (2003). Large capacity storage of integrated objects before change blindness. *Vision Research, 43*, 149–164. [PubMed]
- Levin, D. T., Drivdahl, S. B., Momen, N., & Beck, M. R. (2002). False predictions about the detectability of unexpected visual changes: The role of beliefs about attention, memory, and the continuity of attended objects in causing change blindness. *Consciousness and Cognition, 11*, 507–527. [PubMed]
- Lien, M.-C., Ruthruff, E., Goodin, Z., & Remington, R. W. (2008). Contingent attentional capture by top-down control settings: Converging evidence from event-related potentials. *Journal of Experimental Psychology: Human Perception and Performance, 34*, 509–530. [PubMed]
- Livingstone, M., & Hubel, D. (1984). Anatomy and physiology of a color system in the primate visual cortex. *Journal of Neuroscience, 4*, 309–356. [PubMed] [Article]
- Livingstone, M. S., & Hubel, D. H. (1987). Psychophysical evidence for separate channels for the perception of form, color, movement, and depth. *Journal of Neuroscience, 7*, 3416–3468. [PubMed] [Article]
- Loftus, G. R., & Mackworth, N. H. (1978). Cognitive determinants of fixation location during picture viewing. *Journal of Experimental Psychology: Human Perception and Performance, 4*, 565–572. [PubMed]
- Mitroff, S. R., & Simons, D. J. (2002). Changes are not localized before they are explicitly detected. *Visual Cognition, 9*, 937–968.
- Mitroff, S. R., Simons, D. J., & Levin, D. T. (2004). Nothing compares 2 views: Change blindness can occur despite preserved access to the changed information. *Perception & Psychophysics, 66*, 1268–1281. [PubMed] [Article]
- Mulckhuysen, M., Van Zoest, W., & Theeuwes, J. (2008). Capture of the eyes by relevant and irrelevant onsets. *Experimental Brain Research, 186*, 225–235. [PubMed] [Article]
- Mullen, K. T. (1985). The contrast sensitivity of human colour vision to red-green and blue-yellow chromatic gratings. *The Journal of Physiology, 359*, 381–400. [PubMed] [Article]
- Navalpakkam, V., & Itti, L. (2005). Modeling the influence of task on attention. *Vision Research, 45*, 205–231. [PubMed]
- Neo, G., & Chua, F. K. (2006). Capturing focused attention. *Perception & Psychophysics, 68*, 1286–1296. [PubMed] [Article]
- Noë, A., Pessoa, L., & Thompson, E. (2000). Beyond the grand illusion: What change blindness really teaches us about vision. *Visual Cognition, 7*, 93–106.
- O'Regan, J. K., & Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences, 24*, 939–1011. [PubMed]
- Palmer, S. E. (1975). The effects of contextual scenes on the identification of objects. *Memory and Cognition, 3*, 519–526.



- Parkhurst, D., Law, K., & Niebur, E. (2002). Modeling the role of saliency in the allocation of overt visual attention. *Vision Research*, *42*, 107–123. [PubMed]
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437–442.
- Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, *32*, 3–25.
- Rensink, R. A. (2002). Change detection. *Annual Review of Psychology*, *53*, 245–277. [PubMed]
- Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science*, *8*, 368–373.
- Schreij, D., Owens, C., & Theeuwes, J. (2008). Abrupt onsets capture attention independent of top-down control settings. *Perception & Psychophysics*, *70*, 208–218. [PubMed] [Article]
- Scott-Brown, K. C., Baker, M. R., & Orbach, H. S. (2000). Comparison blindness. *Visual Cognition*, *7*, 253–267.
- Shore, D. I., & Klein, R. M. (2000). The effects of scene inversion on change blindness. *Journal of General Psychology*, *127*, 27–43. [PubMed]
- Simons, D. J., & Levin, D. T. (1997). Change blindness. *Trends in Cognitive Science*, *1*, 261–267.
- Smith, C. N., Hopkins, R. O., & Squire, L. R. (2006). Experience-dependent eye movements, awareness, and hippocampus-dependent memory. *Journal of Neuroscience*, *26*, 11304–11312. [PubMed] [Article]
- Stirk, J. A., & Underwood, G. (2007). Low-level visual saliency does not predict change detection in natural scenes. *Journal of Vision*, *7*(10):3, 1–10, <http://www.journalofvision.org/content/7/10/3>, doi:10.1167/7.10.3. [PubMed] [Article]
- Tatler, B. W., Baddeley, R. J., & Gilchrist, I. D. (2005). Visual correlates of fixation selection: Effects of scale and time. *Vision Research*, *45*, 643–659. [PubMed]
- Torralba, A., Oliva, A., Castelhano, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. *Psychological Review*, *113*, 766–786. [PubMed]
- Treue, S. (2003). Visual attention: The where, what, how and why of saliency. *Current Opinion in Neurobiology*, *13*, 428–432. [PubMed]
- Triesch, J., Ballard, D. H., Hayhoe, M. M., & Sullivan, B. T. (2003). What you see is what you need. *Journal of Vision*, *3*(1):9, 86–94, <http://www.journalofvision.org/content/3/1/9>, doi:10.1167/3.1.9. [PubMed] [Article]
- Underwood, G., & Foulsham, T. (2006). Visual saliency and semantic in congruency influence eye movements when inspecting pictures. *Quarterly Journal of Experimental Psychology*, *59*, 1931–1949.
- Underwood, G., Foulsham, T., van Loon, E., Humphreys, L., & Bloyce, J. (2006). Eye movements during scene inspection: A test of the saliency map hypothesis. *European Journal of Cognitive Psychology*, *18*, 321–342.
- Underwood, G., Humphreys, L., & Cross, E. (2007). Congruency, saliency and gist in the inspection of objects in natural scenes. In R. P. G. van Gompel, M. H. Fischer, W. S. Murray, & R. L. Hill (Eds.), *Eye movements: A window on mind and brain* (pp. 89–110). Oxford: Elsevier.
- Underwood, G., Templeman, E., Lamming, L., & Foulsham, T. (2008). Is attention necessary for object identification? Evidence from eye movements during the inspection of real-world scenes. *Consciousness and Cognition*, *17*, 159–170.
- Verma, M., & McOwan, P. W. (2009). Generating customised experimental stimuli for visual search using genetic algorithms shows evidence for a continuum of search efficiency. *Vision Research*, *49*, 374–382. [PubMed]
- Wright, M. J. (2005). Saliency predicts change detection in pictures of natural scenes. *Spatial Vision*, *18*, 413–430.
- Wright, M. J., Green, A., & Baker, S. (2001). Limitations for change detection in multiple Gabor targets. *Visual Cognition*, *7*, 237–252.
- Yarbus, A. L. (1967). *Eye movements and vision*. New York: Plenum Press.