

# Perceptive Fields of Saliency

Frédéric J.A.M. Poirier

Département de Psychologie,  
Université de Montréal, Canada



Frédéric Gosselin

Département de Psychologie,  
Université de Montréal, Canada



Martin Arguin

Département de Psychologie,  
Université de Montréal, Canada

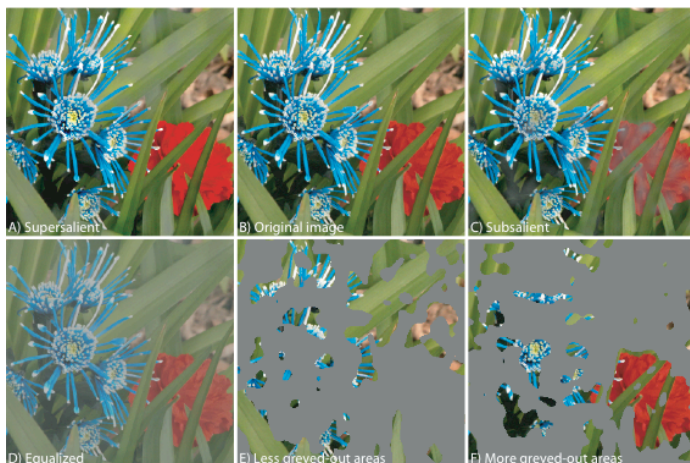


Visual saliency plays an important role in early vision. Humans automatically orient to salient information, via covert attentional shifts and overt eye movements. Here, we measured saliency using a novel psychophysical method. The stimulus was a grid of colored oriented lines. Line luminance varied continuously over the image. Using a mouse, participants adjusted line luminance at locations in the image, until all lines appeared homogeneously luminant. Participants tended to increase (or decrease) luminance of lines where perceptually salient information was absent (or present), thus line luminance setting correlates with perceived saliency. Perceptually less salient image regions were more homogenous in color and orientation, consistent with iso-feature suppression. Perceptual fields of contextual modulation are derived, showing increased saliency surrounding color and/or orientation changes, increased saliency for collinear and end-stopping lines, and a nonlinear integration of saliencies across dimensions. It took 3 or more surround items identical to a target to generate a measurable inhibitory effect, beyond which every additional identical item had a monotonic effect. These novel findings allow a revision of current models of visual saliency. In particular, we found evidence of sustained saliency. Moreover, this new method is sensitive within the normal functioning range, unlike most current research methods.

Keywords: search, attention, computational modeling, perceptual organization, receptive fields

## Introduction

Attention is initially and automatically drawn to salient information. For example, in a natural scene such as shown in [Figure 1](#), we tend to attend to the flowers and not to the leaves. Not surprisingly then, saliency influences performance in many tasks as it guides attention towards (or away from) task-relevant information.



[Figure 1](#). Using the techniques presented in this article, high-saliency areas of a natural scene (B; from Olmos & Kingdom, 2004) can be identified. This information can be used to emphasize saliency differences (A) or to reduce them (C). To identify high-saliency areas in the image, participants adjusted the ratio of image-to-grey locally until the image appeared as homogeneously salient (D; see methods for details). Once adjusted (D), some areas were less greyed-out than others (shown in E and F respectively). It is assumed here that participants grey-out high-saliency areas more than low-saliency areas. In this example, the “equisaliency distance” is defined as the difference between the adjusted image and a control image containing the same amount of grey uniformly distributed. The modified images (A & C) are shown at twice the equisaliency distance from the original image. In the image where saliency differences were emphasized (A), notable differences include increased saturation and luminance of the red flower and the yellow of the blue flowers, at the expense of decreased saturation and contrast of the background leaves.

Several lines of research have contributed to our understanding of saliency, including research on eye-movements, computational modeling, neurophysiology, and psychophysics. The most relevant to the research presented in this article is psychophysical; we will attempt to

bridge our efforts with the other lines of research in the general discussion.

The typical psychophysical approach to measure saliency rests on how it affects performance: cues are said to have high saliency to the degree that they improve (or impair) performance when they are valid (or misleading) by drawing attention to (or away from) the target location (Abrams & Christ, 2005; Folk, Remington & Johnston, 1992; Franconeri, Hollingworth & Simons, 2005; Hillstrom & Yantis, 1994; Huang & Pashler, 2005; Jonides & Yantis, 1988; Kim & Cave, 1999; Koene & Zhaoping, 2007; Nakayama & Mackeben, 1989; Nothdurft, 1993a, 2002; Sobel & Cave, 2002; Theeuwes 1991, 1994; van Zoest & Donk, 2005, 2006; Yantis & Egeth, 1999). Saliency effects occur prior to target identification (Nothdurft, 2002, 2006; Sagi & Julesz, 1985). Research has identified many cues that can modulate saliency hence performance as well (e.g. luminance, color, motion, onset/offset, closure). However, the typical psychophysical approach relies on performance measurements, which usually involve other high-level processes such as memory, task demands, attention, and search strategies, thus include confounding variables.

Nothdurft introduced a “saliency match” task (1993b, 2000a), where instead of relying on performance measurements, he measured directly the perceived saliency of a stimulus. In his experiments, two arrays of items were presented simultaneously, both with identical background elements, except that one array contained a comparison item that differed in luminance (a strong inducer of saliency: Braun, 1994; Nothdurft, 2000b, 2002; but see Einhäuser & König, 2003), whereas the other array contained a target item that differed in some other dimension. Participants indicated which array contained the most salient item, and the luminance of the comparison item was adjusted until comparison and target items appeared equally salient. That is, he measured how “bright” the target item appeared to participants. Using variants of this methodology in a series of experiments, he established that (1) targets defined as singletons in two dimensions were not as salient as expected by the sum of saliencies, suffering a gain reduction effect (Nothdurft, 2000a), (2) luminance suffered less gain reduction than other dimensions (Nothdurft, 2000a), and (3) feature-to-context relations rather than features per se defined saliency (Nothdurft, 1992, 1993a), as shown by context-dependency effects (Nothdurft, 1993b).

In the study reported below, we modified Nothdurft’s saliency matching task, such that luminance can be equated at all locations in an image by adjusting luminance levels at any location within the image. As in Nothdurft’s method, we use luminance as a correlate of saliency. Data collection is more efficient than traditional methods because (1) participants can and usually do respond more frequently than in traditional psychophysical experiments, and (2) each response contains more information than that in a 2AFC task since it indicates the location and level of the perceived saliency. The result is a saliency map with the same extent

as the original image. Analytical tools can then be used to describe how saliency changes with respect to local homogeneity, attribute combinations, collinearity, and distance between elements, as well as to generate perceptive fields of context modulation. Instead of using a natural image as in Figure 1, we present our method on an artificial image composed of colored oriented lines (two colors, two orientations).

## Methods

### Participants

Ten participants volunteered (7 females), including the first and second authors, as well as university undergraduate and graduate students. Their vision was normal or corrected to normal. Participants were paid 10\$/hour.

### Apparatus

Testing and data collection was done on a PC computer (P4 3GHz) set to a resolution of 800x600 pixels and a refresh rate of 75 Hz. Responses were recorded via mouse button presses. Viewing distance was 68.5cm, where each item occupied a 0.5° x 0.5° square (8 x 8 pixels), and the entire stimulus subtended a 16° x 16° area.

## Procedure

### Visual spread task

Participants were presented with a display containing a grid of items, each item was a red or green right- or left-oblique line over a black background. Image luminance was varied over the image from 0% (black) to 100% (image seen clearly; see below for details). Participants were required to equate apparent luminance over the image by left-clicking (or right-clicking) on points where luminance was perceived as higher (or lower) than elsewhere, which decreased (or increased) the luminance within a small window around that location. The position and duration of button presses was used to dynamically adjust luminance (see below for details).

**Stimuli.** Stimuli were made of two components: (1) the signal containing the orientation and color information (see below), and (2) the selection field, which controlled the local luminance of the signal (see Figure 2; see below). Stimuli were the pointwise product of the two.

The *signal* was defined as a 32 cells x 32 cells grid of items. An item was a  $\pm 45^\circ$  line subtending an 8x8 pixel area (0.5x0.5° of visual angle), where luminance was 100% on the diagonal, and 73% for adjacent pixels to reduce aliasing. The lines were red or green, with the corresponding RGB value set to its maximum, and the other values set to 0. Calibration is discussed below.

In order to increase the similarity of adjacent items, spatial correlations were introduced in the signal. This was done such that the stimulus contained some homogeneous regions and some heterogeneous regions. Each item was assigned a random value, and those values were spatially blurred using a Gaussian function (G):

$$G_{i,j} = e^{-\frac{(i^2+j^2)}{\sigma^2}} \quad (1)$$

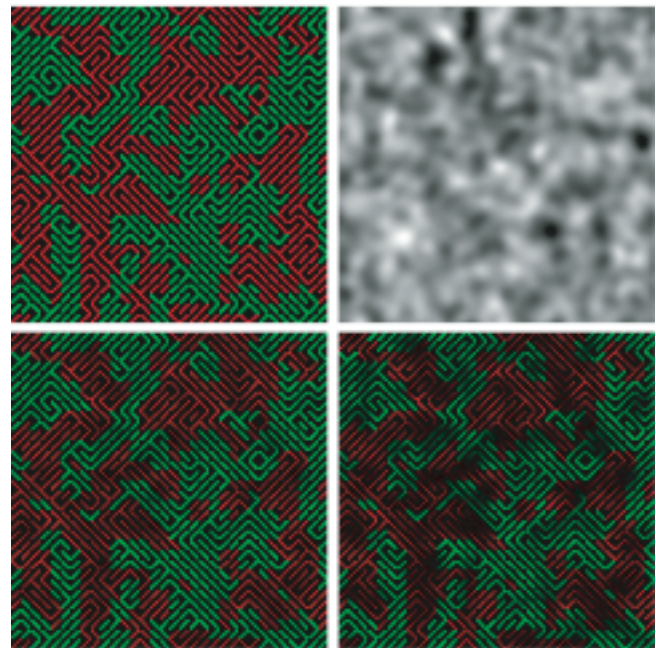
where  $\sigma=1.4$  items, and coordinates (i,j) are in items. The cutoff value was set at the mean, with items above the mean colored red, and all other items colored green. The same steps were used independently to determine item orientation. Color and orientation were independent of each other ( $r = 0.01$ ).

*Selection fields* were used to determine the local intensity of the signal stimulus described above, i.e. the RGB values of the signal above were multiplied by the percentage values from the selection fields to determine the RGB values used in the stimulus displayed. The percentage affected the luminance of lines, but did not affect the luminance of the black background. As such, the local contrast was also affected, but for simplicity, we will refer to luminance changes. For each participant, 10 selection fields were created independently at the onset of the experiment. Selection fields were generated by filtering a 256x256 pixel binary field smoothed using a Gaussian filter ( $\sigma=0.5^\circ$ ), and scaled to a range of intensities between 25% and 75%. During the experiment, selection fields were always normalized to an average of 50%, with values outside the range of 0%-100% clipped. The normalization to an average of 50% was done to prevent participants from setting luminance at ceiling or floor values, as well as to keep average luminance constant across participants and selection fields.

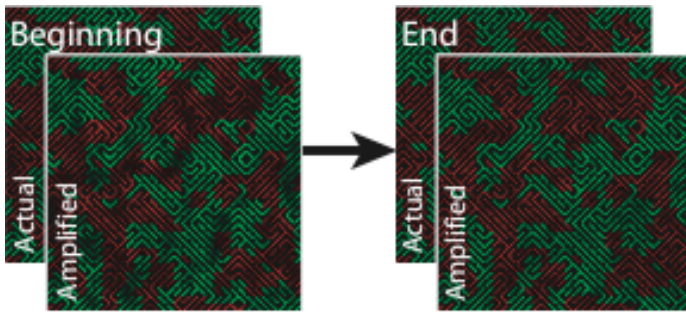
On each trial, a selection field was selected and pointwise multiplied with the signal to create the stimulus. On 50% of trials, the selection field was selected randomly. On the other 50% of trials, the selection field that had least been presented so far was selected, such that all selection fields were presented about equally often throughout the experiment. Changes made to a selection field were kept for future trials, that is, the adjustments made were cumulative. Also, although there were 10 selection fields per participant, there was only one signal used for all participants. The signal and selection field were always aligned before being combined.

Throughout the experiment, stimulus items were presented well above contrast threshold. Thus, they always re-

mained clearly visible. Moreover, the attributes themselves were also at high contrast, with  $90^\circ$  difference in orientation, and colors of opposite polarity and high saturation. Also, the participant's task was to equate luminance across the display, which effectively means that luminance and contrast becomes more homogeneous over the image as participants made more responses. That is, the stimuli displayed to participants towards the end of even the first session appeared fairly homogeneous in luminance and contrast, and the selection fields analyzed below contained relatively little luminance and contrast variance. These values were chosen to maximize the sensitivity of the experiment to the effects of saliency.



**Figure 2.** Stimulus design. (Top left) The signal used to construct the stimulus was an array of red- or green-coloured oblique lines. A “selection field” (top right) was used to modulate the luminance of the lines locally, and could vary between 0% and 100%. The signal and the selection field were combined using pointwise multiplication, that is, the selection field controlled what percent of the signal's luminance was shown at every location. The resulting stimulus (bottom left) looks like the signal, except for a variable luminance. The differences shown in the bottom left image are within the range of luminance amplitudes used at the beginning of the experiment. They are also shown emphasized in the bottom right image. See text for details.



**Figure 3.** During the experiment, participants adjusted local luminance to make it more homogeneous throughout the image. They did so by moving the mouse to a location that they perceived as deviating from apparent equiluminance, and pressed one of two mouse buttons to increase or decrease the luminance within a small Gaussian window around that location. Thus, perceived luminance was inhomogeneous initially (left), but through the course of the experiment became increasingly homogeneous (right). Despite appearing as homogeneous in luminance, the end image contained luminance deviations which can be made apparent either by amplifying them, or through analyses. See text for details.

### Temporal sequence and analysis

To prevent local adaptation and stimulus-border effects, the stimulus was spatially shifted randomly on every trial using tiling. In other words, (1) four copies of the stimulus were tiled using translational symmetry into a square composite, (2) from that composite was sampled a square stimulus of equal size as the original stimulus, with random location within the composite, and (3) the sampled stimulus was presented during the trial, centered on the screen.

During a trial, participants could freely inspect the stimulus and generate as many responses as they wished. The selection field was updated in real-time using the participant's responses, by adding or subtracting a Gaussian (Equation 1,  $\sigma=0.5^\circ$ , step size = 8 RGB steps) at the location of the click for every screen refresh that the button remained pressed. Trials normally terminated after 20 seconds. To prevent trial termination during the participant's response, the current trial was extended to 2 seconds after the participant's last response. At the end of the trial, a blank screen replaced the stimulus for 500ms, and the next trial was then initiated. Testing was conducted in three sessions: one initial session of one hour, and two additional sessions of 15 minutes each, regardless of the number of trials completed. Each subsequent testing session started with selection field values taken from the end of the previous session. This is equivalent to providing more time for participants to make adjustments on the same set of 10 selection fields.

Only line luminance was adjusted during the experiment and analyzed in the results section, as background luminance was unaffected by selection fields. As such, pixels belonging to the black background were excluded from

all analyses. RGB values as adjusted by participants were then converted to luminance values prior to analyses. The conversion use the best-fit 2<sup>nd</sup> degree polynomial relating RGB values and measured monitor luminance values:

$$L_R = 5.46 + 28.53(R/255)^{2.11}, \text{ and} \quad (2)$$

$$L_G = 5.14 + 87.10(G/255)^{2.16} \quad (3)$$

where R and G are RGB values for red and green pixel respectively (range varied between 0 and 255; B is always equal to 0), and  $L_R$  and  $L_G$  are the luminance values for red and green pixels respectively.

All analyses below are corrected for color equiluminance by the following procedure: (1) the average and standard deviation of luminance was calculated over three sets of pixels; which are: (a) all pixels belonging to lines, (r) pixels belonging to red lines only, and (g) pixels belonging to green lines only, (2) the luminance of (r) and (g) were normalized such that their averages and standard deviations were equal to those of (a). This was done for each participant separately. Additional analyses on raw RGB values and non-normalized luminance values gave similar results to those reported below, provided that results were averaged across colors.

## Results

Data was averaged across the 10 selection fields per participant, and analyses were performed per participant. The results of these analyses were then combined across participants to generate the means and error bars (standard error of the mean) shown in figures, using the 10 participants as samples. [] The error bars thus represent a measure of inter-subject variability.

Three testing sessions were included, where subsequent sessions continued where the previous session ended (see details above). The last two sessions improved statistical power and reliability, but otherwise did not change the nature of the effects. Therefore, we present only the results after completion of the third session, but note that similar results were found after the first session.

Participants systematically set the local luminance higher in certain image locations and lower in others, thus resulting in systematic luminance modulations over the image (see Figure 4). Following the logic that salient parts of the image were perceived as brighter, participants would set those points at a lower luminance to perceive them as equally bright as other parts of the image. In other words, the point of subjective equality in luminance would be negatively correlated with perceived saliency. We henceforth use the term "saliency" to denote differences between average luminance and local luminance (i.e. saliency = average luminance over the image minus local luminance), which is measured in  $\text{cd/m}^2$ . For the purposes of this article, saliency is thus measured in units of luminance in a

manner similar to that proposed by Northdurft (see Introduction). Note that the negative relationship between luminance and saliency in our experiment is a natural consequence of having “saliency due to luminance” (as adjusted by participants) compensate for “saliency due to other factors” (as defined by stimulus characteristics), and is consistent with the more usual definition of saliency where perceived saliency generally increases with increased luminance and contrast.

## Surface analysis

Using the saliency measure, the stimulus was broken down into regions of various saliency levels, using the following procedure. Each item of the stimulus was classified as belonging to one or more of five saliency categories in 2 steps. (1) Pixels were classified into five saliency categories (pixels belonging to the two extreme categories are shown in the left half of Figure 4), ranging from suppressed (highest luminance range) to neutral (middle luminance range) and to salient (lowest luminance range). The four cutoff values for those ranges were set at the average luminance value  $\pm 1/6^{\text{th}}$  and  $\pm 1/20^{\text{th}}$  of the maximum deviation from the average luminance value. These cutoff values were thus chosen to spread the data about equally between the various categories. However, other cutoff values yield similar results. (2) Items were included into saliency categories if at least one pixel of that category overlapped the item. Thus items could belong to two or more categories at once. Allowing items to belong to more than one category improved the sensitivity of the analysis, but otherwise did not change the pattern of results.

Once items were classified into saliency categories, homogeneity in color was calculated using the following procedure for each saliency category separately. (1) Items that do not belong to the chosen saliency category (e.g. salient, neutral, or suppressed) were removed from further analysis.

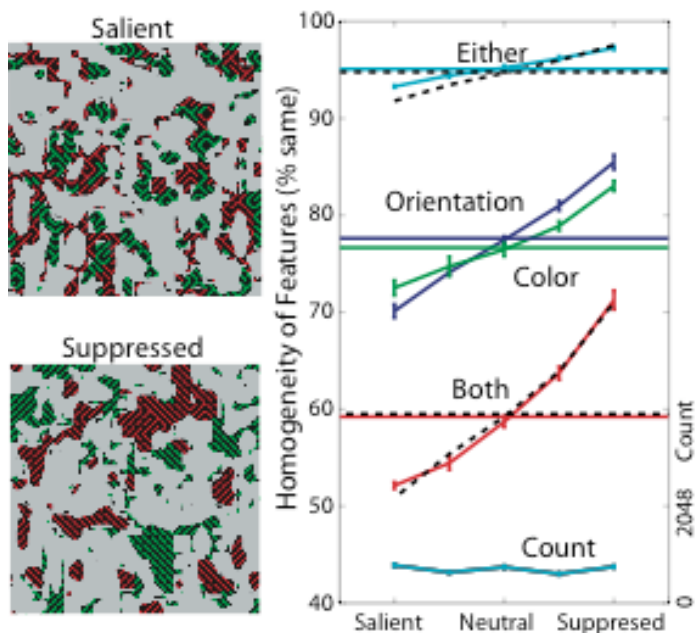
(2) The proportion of items adjacent to the target item (i.e. items offset by  $0.5^\circ$  from each other either horizontally or vertically) of the same color was calculated over the remaining items. The same analysis was repeated for homogeneity in orientation, and specific combinations of color and orientation (see below).

There is a clear relationship between saliency and homogeneity, where saliency decreases as homogeneity increases (see Figure 4), which is consistent with iso-feature surround suppression mechanisms (Blasdel, 1992; Das & Gilbert, 1999; Knierim & Van Essen, 1992; Li, 1998, 1999, 2002; Pettet & Gilbert, 1992; Stettler, Das, Bennett & Gilbert, 2002). This effect was found with (1) color homogeneity independently of orientation, (2) orientation homogeneity independently of color, (3) for items that are identical on both dimensions (“both”), and (4) for items that are identical in at least one dimension (“either”). The homogeneity of feature combinations (i.e. both, either) was well-predicted by simple probabilistic combinations of homogeneities of the features themselves (i.e. color, orientation), for all categories of saliency (see dashed lines on Figure 4). Specifically, homogeneities were predicted by joint probabilities, assuming independence between the two attributes, and using the complementary in Equation 4:

$$H_{\text{Either}} = 1 - (1 - H_{\text{Color}}) * (1 - H_{\text{Orientation}}), \text{ and} \quad (4)$$

$$H_{\text{Both}} = H_{\text{Color}} * H_{\text{Orientation}} \quad (5)$$

where H is the homogeneity, expressed as the probability of adjacent items identical to the target item on the relevant dimension. This suggests an absence of strong nonlinearities in the combination of saliency across dimensions (see below for more complete analyses). These effects were found in all participants.

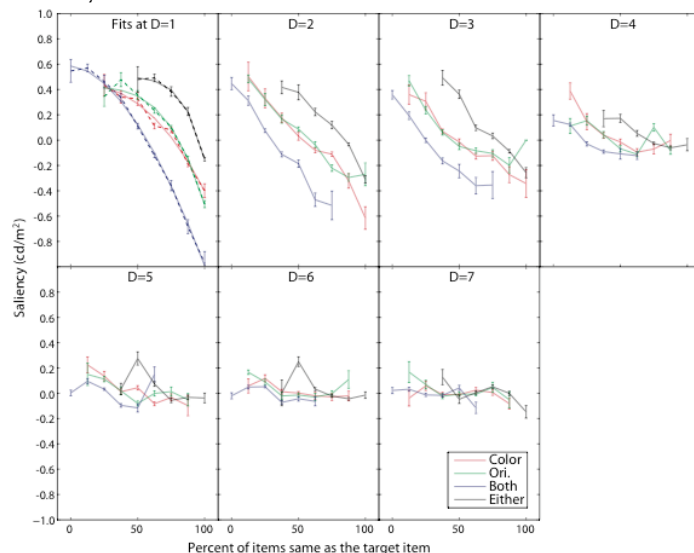


**Figure 4.** Heterogeneous surfaces are more salient than homogeneous surfaces. For this analysis, each pixel was classified into one of 5 saliency categories based on average luminance settings. (Left) These two images show pixels that were categorized as high-saliency (top left; “salient”) or low-saliency (bottom left; “suppressed”). (Right) For each saliency category, feature homogeneity was measured over all pairs of vertically or horizontally adjacent items, including only items that contained at least one pixel of that saliency category. Cutoff values were selected such that each category contained about the same number of item pairs for analysis. Homogeneity was measured as the proportion of adjacent items that shared the feature (i.e. color or orientation; error bars = S.E.M.). Horizontal bars show the baseline homogeneity for the whole stimulus. “Both” and “either” refer to the proportion of item pairs that were identical on both or either dimension respectively. Dashed lines represent expected homogeneities of “both” and “either” predicted from combining homogeneities of color and orientation. Count shows the number of item pairs included in the saliency category. There was a strong relationship between homogeneity (Y axis) and saliency (X axis) for every feature and combination of features analysed. See text for details.

### Simple predictors of saliency

Figure 5 shows the average saliency of items (X axis) as a function of the percentage of surrounding same-color or parallel items, and as a function of the distance. These surround items were sampled at distances (measured as the number of items away from the target item; D in Figure 5). “Percent of items same as the target item” is based on 8 sampled items per target item (4 sampled at cardinal directions, and 4 sampled at diagonal directions at the nearest integer distance). Only averages of at least 10 occurrences were included. Singletons (i.e. unique items embedded in otherwise homogeneous fields) are not shown due to their

low frequency of occurrence. At small distances, there is a clear effect of the percentage of identical items. This effect is much smaller at distances of about 4-5 items for color, and 3-4 items for orientation. There is no indication that image statistics beyond 5 items can be used to predict local saliency.



**Figure 5.** An item’s saliency is higher if nearby items have a different color and/or orientation. The average saliency (Y axis) of a target item is shown as a function of the percent of items surrounding it (X axis) that are identical to it in color (red), orientation (green), both properties (blue) or either property (black). The surrounding items used in this analysis were sampled in 8 directions at given distances (D in items; diagonal distance rounded to the nearest integer). Saliency was higher when nearby items (i.e. low D values) were different in color and/or orientation (i.e. low percent values on the X axis). Also shown at the closest distance (D=1) are power-function fits to the data (dashed lines for data, solid lines for fits). See text for details.

The most interesting result is for nearby items (distance or D = 1) because there are no items between the target and the surround items that may interfere with their interaction. At this distance, for both color and orientation, saliency is approximately invariant when the target shares a property (color or orientation) with 25%-50% of surrounding items, and sharply drops as similarity increases beyond that. Similar types of functions are found for homogeneity in feature combinations (i.e. both, either). This data may be summarized by the general rule that decreasing target-surround similarity increases saliency, and that increase in saliency is larger when target-surround similarity is high than when it is low. That is, the target item’s saliency (S) decreases as more surround items are similar to the target item, as approximated by the power function:

$$S = A - B(N_A / N_{max})^C \tag{6}$$

where  $N_A$  is the number of adjacent elements surrounding the element of interest that share the feature or feature combination (e.g. color, orientation, both, either),  $N_{max}$  is

the maximum possible number of adjacent elements ( $N_{\max} = 8$ ), and the constants A, B, and C were best-fit using the `fminsearch` search function from the Matlab Optimization toolbox (MathWorks Ltd.) (fits are shown in Figure 5, top-left panel;  $R^2$ 's were 95%-99.7% across features and feature combinations; see Table 1). The constant C was always greater than unity (color: 2.39, orientation: 3.73, both: 1.77, either: 6.41). In other words, it takes at least 3 identical items (or 5 items with at least one identical dimension) for inhibition to start having a measurable effect, beyond which every additional identical item has a monotonic effect.

Feature category	A	B	C	$R^2$
Color	0.4505	0.8506	2.3946	96.17%
Orientation	0.4187	0.9272	3.7264	95.00%
Both	0.5859	1.5770	1.7657	99.74%
Either	0.5008	0.6514	6.4067	98.09%

Table 1: N Effects (see Equation 6; see Figure 5)

The dependence of saliency on target-surround similarity is often believed to arise from iso-feature suppression (Li, 1999), which is usually modeled as either linear inhibition or divisive inhibition. Linear inhibition postulates that each similar surround item subtracts a constant quantity of saliency, and would predict that the constant C in Equation 6 would be 1. Divisive inhibition postulates that the target item's saliency is divided by each similar surround item, in which case the best-fit to that function using Equation 6 would give a constant C between 0 and 1. That is, both linear inhibition and divisive inhibition would provide lower-quality fits to the data shown in Figure 5 top left panel, because they both incorrectly predict the first few identical items should decrease saliency at least as much as additional items.

Note that background homogeneity in our experiment is equal to  $|100\% \cdot N_A / N_{\max} - 50\%| + 50\%$ . Background homogeneity is high at both ends of the target-background similarity continuum in our experiment (i.e. background homogeneity is 100% when  $N_A / N_{\max} = 0\%$  or 100%). It is thus clear that background homogeneity alone cannot account for the results. Our results are thus dependent on target-background similarity.

## Saliency Perceptive fields

We generated perceptive fields of context modulation of saliency (shown in Figure 6) using techniques similar to those used to generate Figure 5. By analogy, we measured how the perceived saliency of a target item placed at the center of a receptive field (i.e. the target item) changed as a function of a surround item's relative position and similarity to the target item. Note that whether we measure the influence of a surround item on the saliency of a center item or *vice-versa* has no influence on the results, because our analysis assumes reciprocal interactions. That is, our analysis was not sensitive to possible asymmetries regarding

how different feature types may interact with one another (e.g. Foster & Ward, 1991; Poirier & Gurnsey, 1998; Sagi & Julesz, 1985; Treisman & Gormican, 1988; Treisman & Souther, 1985).

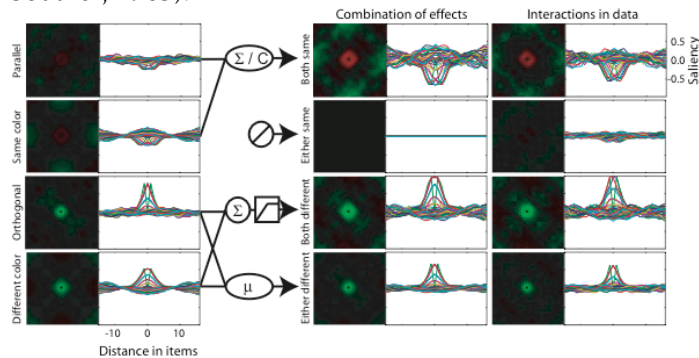


Figure 6. Perceptive fields of saliency. These graphs show how the saliency of an item is influenced by the similarity and relative position of another item. Each graph is shown twice, as a 2D plot to emphasize spatial scale, and as a line plot to emphasize amplitude. In the line plots, each line corresponds to a row from the 2D plot. The 2D plot shows increases (green) or decreases (red) in average item saliency, compared to average saliency (black). Each pixel represents one item in the stimulus, and the central black pixel represents the central item serving as basis for comparison. Thus the perceptive fields shown are the same size as the stimulus. Whenever orientation was a factor, the central item used was a right oblique (after transformations). (Left column) From top to bottom, perceptive fields of saliency are shown for within-dimension comparisons, namely: (1) parallel items, (2) same-color items, (3) orthogonal items, and (4) different-color items. For example, the "Different color" graph shows an increase in saliency for nearby items that differ in color, regardless of direction. (Right column) From top to bottom, perceptive fields of saliency are shown for between-dimension comparisons, namely: (1) "both same", i.e. items that have the same color and orientation, (2) "either same", i.e. items that have the same color and/or orientation, (3) "both different", i.e. items that differ in both color and orientation, and (4) "either different", i.e. items that differ in color and/or orientation. (Middle column) Between-dimension effects can be accounted for as simple combinations of within-dimension effects. From top to bottom, the between-attribute effects were predicted using: (1) a weighted sum of parallel and same-color, (2) no effect, (3) a sum of orthogonal and different color, followed by a by a compressive nonlinearity, and (4) an average of orthogonal and different-color. See Supplementary Figure 1 for individual-participant data. See text for details.

Perceptive fields can be derived for any center-surround relationship (e.g. same color, different orientation, same color and orientation) in 3 general steps (see example below for more details). (1) The central (and surround) item criterion defines which items are included in the analysis and placed in the center (and surround). (2) The image was then shifted and rotated such that in turn, every item that satisfies the central item criterion was placed in the center, thus setting the relative position of surround items. (3)

Saliency was then averaged over every center-surround item pairs included in the analysis, as a function of relative position. Thus we can measure how saliency is dependent on similarity, distance, and alignment (where applicable).

For example, the perceptive field for “different color”, shown in Figure 6 bottom left, was derived as follows: (1) find an item that satisfies the central item criteria (e.g. red), (2) align the signal and the saliency data such that this item is centered, (3) sum the saliency scores for every item satisfying the surround criteria (e.g. green) at those locations, (4) repeat steps 1-3 for every item satisfying the central criteria, (5) divide by the number of items satisfying these criteria to obtain the average saliency score, (6) repeat steps 1-5 for different criteria of interest (e.g. red in green, and green in red), and (7) assuming no directional bias (e.g. left vs. right, above vs. below), congruently with computational models of saliency, average the results of step #6 over mirror-images and 90°-rotations of itself. The results of that computation can be displayed to emphasize spatial relationships and effect amplitudes (e.g. Figure 6 bottom left). Similar analyses were performed on orientation and various combinations of color and orientation, except that analyses including orientation only included mirror-images and 90°-rotations that produced a right-oblique central item, to preserve collinear and flanking relationships (i.e. whether two parallel items are aligned or side-by-side).

The resulting perceptive field for color differences (see Figure 6 bottom left) shows how the average saliency is modulated when color changes over given locations. Specifically, two differently-colored items increase each other’s saliency over a short range. In contrast, saliency modulation is much weaker between same-color items (Figure 6 left, 2<sup>nd</sup> row). These results are consistent across participants (see Supplementary Figure 1).

The same technique used to generate perceptive fields for color (see above) was applied to orientation and combinations of orientation and color. Figure 6 (left, rows 1 & 3) present data for identical and different item orientations (i.e. parallel and orthogonal). The data was normalized to a right-oblique central item in order to emphasize collinearity and flanking effects. Again, it is clear from Figure 6 (left, row 3) that saliency increases for orthogonal items, over a short range. Items are inhibited when parallel items are located anywhere around it (Figure 6, left, row 1). Further analyses also revealed more subtle effects: (1) collinear items suffered less inhibition than flanking items, which was observed both independently of color and when color was identical ( $t$ 's (9) = 1.82, 1.78,  $p$ 's = 0.051, 0.054 respectively), and (2) *end-stopping* lines (i.e. lines placed at the end of other lines at orthogonal orientations) were more salient than *end-stopped* lines (i.e. lines placed on either side of other lines at orthogonal orientations), which was observed both independently of color, and when color also changed ( $t$ 's (9) = 1.46, 2.53,  $p$ 's = 0.09, 0.016 respectively). To sum up, a line placed near the end of another line will be slightly more salient than if it is placed on either side.

Above we derived perceptual fields of context modulation for single attributes and their combinations. These perceptual fields are related to each other. Indeed, each of the four combinations investigated (shown in Figure 6 right) can be approximated by a combination of single attribute effects (combination rules and predictions shown in Figure 6 middle;  $R^2 = 77.8\%$  across all 4 interactions). When items were the *same* (i.e. identical in both attributes), perceptive fields of saliency were well described by a sum of the individual effects, except for a gain increase ( $R^2 = 76.7\%$ ). When items were the same on *either* (or both) dimension, saliency effects cancelled each other out. When items were different on at least one dimension (i.e. either different), perceptive fields of saliency were well described by the average of the two individual effects ( $R^2 = 73.7\%$ ). Finally, when items were different on *both* dimensions, perceptive fields of saliency were well predicted by the sum of the two effects followed by a compressive nonlinearity ( $R^2 = 93.0\%$ ).

Overall, saliency increases due to feature differences have greater amplitudes and narrower spatial extents than saliency decreases due to feature similarities.

## Models

### Overview

We account for the data reported above in this article using two models. The first model uses the perceptive fields described in the last section (i.e. Saliency Perceptive fields) to predict human data. The second is based on a published model of visual search (Wolfe, 1994), with some free parameters added to improve the quality of the fit. In both cases, the goal is to assess how much of the data can be explained using simple mechanisms.

For both models, the free parameters were adjusted with the goal of decreasing the sum of squared differences between the human and model data, using the `fminsearch` function in the Matlab Optimization toolbox. The fit was performed on data averaged over all participants.

### Perceptive fields

The stimulus was built using two 32x32 binary maps, one per dimension (e.g. color and orientation), with values indicating if the item was red vs. green and left- vs. right-oblique. These binary maps, henceforth called search maps, were convolved by the perceptive fields extracted above (see the 8 perceptive fields in Figure 6, not including the middle column), and saliency was defined as a weighted sum using an exponent for compression:

$$S' = \omega_0 + \sum \omega_i |M_i * PF_i|^\alpha \quad (7)$$

where  $PF_i$  represents the perceptive fields derived from the data, and  $M_i$  represents the associated search maps. The



free parameters included in the fit include a compression exponent ( $\alpha$ ), and weights ( $w_0, w_i$ ; where  $i=1..8$ ).

This analysis is inherently circular: the perceptive fields that have been derived from data are then used to account for variance in that same data. It remains relevant nevertheless because it provides a measure of the variance captured by the perceptive fields. Specifically, it measures how much of the variance in the data is accounted for by the compact representation provided by the perceptive fields derived above.

## Guided Search 2.0

The data was also fit using a model inspired by the Guided Search 2.0 model (Wolfe, 1994), to which we added a component for collinearity and an upper limit to saliency after combination across attributes. Other more complex computational models (e.g. Itti, Koch & Niebur, 1998; Itti & Koch, 2000, 2001; Itti, 2006; Li, 1998, 1999, 2002) use fairly similar mechanisms built into fairly similar architectures. For example, all these models use concentric iso-suppressive context-modulation, and some of these models also use facilitation by collinearity. The main differences between models refer to the combination rules and normalization issues. Normalization issues are not a concern here because orientation and color contrasts were high. We used a linear combination to simplify the fit, and choose to discuss combination rules later (see “Implications for Computational Modelling” section in the Discussion). We thus expect that the results would be similar across different models.

The model used concentric perceptive fields (PF) described as:

$$PF = \frac{e^{-\frac{(x^2+y^2)}{0.434}} - 0.11e^{-\frac{(x^2+y^2)}{1.74}}}{1.783} \quad (8)$$

where  $X$  and  $Y$  are horizontal and vertical distances (in number of items) relative to the center of the perceptive field, respectively. The parameters were chosen to provide a small surround effect, and such that the perceptive field sums to 0 (i.e. gives a response of 0 on homogeneous fields), and the area under the curve sums to 1. Predicted saliency ( $S'$ ) is given as:

$$S' = \omega_0 + \min(S_{summed}, \omega_{Cutoff}), \text{ where} \quad (9)$$

$$S_{summed} = \omega_{Color} |M_{Color} * PF_{Color}|^\alpha + \omega_{Ori} |M_{Ori} * PF_{Ori}|^\alpha + \omega_{Coll} N_{Coll} \quad (10)$$

and  $M_{Color}$  and  $M_{Ori}$  are the feature maps for color and orientation respectively,  $PF_{Color}$  and  $PF_{Ori}$  are the corresponding concentric perceptive fields, and  $N_{Coll}$  is the collinear term indicating the number of adjacent collinear items (0 to 2). The free parameters included in the fit were parame-

ters controlling the weights of different effects (i.e.  $\omega_0, \omega_{color}, \omega_{ori}, \omega_{coll}$ ), a weight controlling compression ( $\alpha$ ), and a weight imposing an upper limit to saliency ( $\omega_{cutoff}$ ; consistent with Wolfe, 1994).

## Modeling comparisons

Although there are differences in the details of the computations performed, the two models above have similar main features. They both use filtering using roughly concentric perceptive fields, include collinear facilitation, combine the effects additively (i.e. as a weighted sum; see Equation 7 & Equation 9), and show compression at higher levels of saliency ( $\alpha$  was 0.67 and 0.78 respectively; and  $\omega_{cutoff}$  was low enough to have an effect on the quality of the fits).

Although the data-driven perceptive fields were similar to the concentric filters used in the Guided Search model, there were some differences. The perceptive field model also includes perceptive fields derived for between-attribute effects, and end-stopping facilitation. The Guided Search model allowed the collinearity weight ( $\omega_{coll}$ ) to vary independently of suppression due to parallelism ( $\omega_{ori}$ ). Moreover, the Guided Search includes a parameter for a ceiling saliency value ( $\omega_{cutoff}$ ), which was not included in our model.

Either model accounts for about half of the variance in the data (51.8% and 54.0% respectively of the variance remaining after compensation for color equiluminance). It is difficult to judge the quality of fits without some proper standard (in this respect, the target of 100% explained variance is unrealistic, as explained below). To provide such a standard, we estimated how much of the variance was reliable across participants. We split our 10 participants in two groups of 5 participants each, averaged the data for each group, and measured the correlation of item saliency across the two groups. Repeated over all possible ways of splitting the participants in two groups of 5, the average  $R^2$  was  $37.6\% \pm 0.5\%$ . In other words, 37.6% of the variance in the data averaged over 5 participants is reliable. From this measure, we can extrapolate using the Spearman-Brown formula (Brown, 1910; Spearman, 1910) that in the average of 10 participants,  $57.3\% \pm 0.5\%$  of the variance is reliable. That is, it appears that either model accounts for about as much variance as we can reliably measure using 10 participants. Note that these estimates represent the reliability of saliency measured per item, whereas the perceptive fields derived above are more reliable due to averaging over items.

The Guided Search model gave equal weights to orientation and color ( $\omega_{ori} = 0.0765$  and  $\omega_{color} = 0.0731$  respectively), but a smaller weight for collinearity ( $\omega_{coll} = 0.00086$ ). The parameters of the perceptive fields model will not be discussed, mainly because it used a combination of linear filters to model nonlinear interactions. This was useful in order to estimate the variance accounted for by the perceptive fields derived. However, we believe that it is best to

delay an interpretation of model parameters until a nonlinear model is developed.

## Discussion

### Summary of findings

Within 1 hour of data collection per participant, this novel and intuitive method produced reliable data on perceived saliency throughout an image. An additional 30 minutes of data collection per participant improved statistical power and reliability, but otherwise did not change the nature of the effects.

Saliency increased with heterogeneity. A target's saliency was lowest when surrounded by identical items, rapidly increased when a few surrounding items were different, and finally reached a plateau at high saliency when most surrounding items were different. This increase was consistent with a power function rather than either linear or divisive inhibition. These effects were found with respect to orientation, color, and their combinations.

Even though cross-attribute effects were well-predicted from single attribute effects, more sensitive analyses revealed reliable deviations from independence. In particular, items identical in orientation and color inhibited each other more than expected by levels of inhibition measured on the separate dimensions, and items different in orientation and color facilitated each other less than expected by levels of facilitation measured on separate dimensions. This is reminiscent of the power-function-like inhibition discussed above (see Equation 6), as in both cases the inhibitory power accelerates as similarity increases.

Both parallel and orthogonal items are more salient when placed at the end of lines than when placed on either side of lines. This increases the saliency of object contours, because they tend to be continuous and terminate internal structure (e.g. internal textures and internal contours).

### Implications for neurophysiology

There are many competing proposals for the location of the saliency map, including the pulvinar (Posner & Petersen, 1990; Robinson & Petersen, 1992; Rockland, Andresen, Cowie & Robinson, 1999), the superior colliculus (Horwitz & Newsome, 1999; Kustov & Robinson, 1996; McPeck & Keller, 2002; Posner & Petersen, 1990), the primary visual cortex (Li, 1998, 1999, 2002; Zhaoping & May, 2007), the frontal eye field (Thompson, Bichot & Schall, 1997; Schall 2002), and the lateral intraparietal area (Colby & Goldberg, 1999; Gottlieb, Kusunoki & Goldberg, 1998). Some researchers argue that saliency is distributed in the brain. The issue is complicated by the use of different definitions of saliency and attention, as well as the reliance on performance measurements in behavioral studies of saliency (Blaser, Pylyshyn & Holcombe, 2000; McAdams & Maunsell, 2000; O'Craven, Downing &

Kanwisher, 1999; Reynolds, Alborzian & Stoner, 2003; Saenz, Buracas & Boynton, 2002; Treue & Martinez-Trujillo, 1999; for reviews, see Assad 2003, and Treue 2001, 2003). The current study can shed some light with regards to stimulus-driven saliency.

Consensus is building that stimulus-driven saliency occurs mainly via iso-feature suppression mechanisms. In the orientation domain, iso-feature suppression is proposed to occur via corticocortical long-range interactions among orientation-tuned units with non-overlapping receptive fields (Blasdel, 1992; Das & Gilbert, 1999; Pettet & Gilbert, 1992; Stettler et al., 2002). Similar iso-feature suppression mechanisms are known to exist in the color domain. Our data is consistent with iso-feature suppression in both domains. In the orientation domain, we also provided evidence supporting the role of both collinear and end-stopping mechanisms in modulating saliency, whereas previous studies have focused on the role of collinear mechanisms (Hubel & Wiesel, 1965; Kapadia, Ito, Gilbert & Westheimer, 1995; Jingling & Zhaoping, 2008; Nelson & Frost, 1985; Polat & Bonneh 2000; Polat, Mizobe, Pettet, Kasamatsu & Norcia, 1998; Polat & Norcia, 1996, 1998; Polat & Sagi, 1993, 1994a, 1994b). The contributions of collinearity and end-stopping do implicate that V1 contributes to saliency, as both of these functions first appear in V1. However, this does not rule out the possibility that this information is relayed to other brain areas where saliency would be computed.

Our data suggests that saliency is more strongly modulated by surround items that are dissimilar rather than similar. Iso-feature inhibition could account for this asymmetry, with a neural implementation consistent with Equation 6. One biologically-plausible neural implementation would be a neuron (1) that sums the activity of neurons with similar preferences over some cortical area, (2) whose output is a power-function of its summed inputs, and (3) whose output inhibits the neuron signaling saliency. This for example could be achieved by inhibitory interneurons in V1 (e.g. Gilbert, 1992; Knierim & Van Essen, 1992; Rockland & Lund, 1983; Wachtler, Sejnowski & Albright, 2003).

### Implications for computational modelling

Several research groups have modeled saliency (e.g. Itti & Koch, 2000, 2001; Li, 1998, 1999, 2002; Tsotsos et al., 1995), with efforts to incorporate knowledge from physiology, in order to account for human performance (Lee, Itti, Koch & Braun, 1999; Itti et al., 1998; Itti & Koch, 2000; Wolfe, 1994; Li, 2002), and predict preferred saccade locations (Carmi & Itti 2006; Itti, 2006; Peters, Iyer, Itti & Koch, 2005; Wolfe, 1994). Common features of these models include: (1) feature-extraction for basic visual cues (e.g. color, luminance, motion, texture, orientation), often abstracted rather than actually extracted from the image, (2) context-modulation by iso-feature inhibition, independently for each feature, (3) a rule for combining saliencies across different features, and (4) mechanisms to generate

behavioral predictions from the saliency map, including saccade locations, reaction time, and response accuracy. The results reported here are relevant to context-modulation (2) and the combination rule (3).

Regarding context-modulation, we have found evidence for iso-feature inhibition (or conversely feature-difference enhancement), which is a common feature of current models. However, we also found evidence for enhancement of collinear and end-stopping line elements. This enhancement, even though somewhat weak, could improve image segmentation by increasing the saliency of line segments belonging to object contours, and suppressing line segments belonging to internal textures. This could help improving the performance of models by increasing the saliency of closed shapes, as well as improving orienting towards objects.

Regarding the combination rule across attributes, the presence of nonlinearities in our data makes it difficult to assess the independence of attributes. As discussed above, incremental changes in homogeneity had a greater effect on saliency when homogeneity was high than when it was low. This effect was observed both within- and between-dimensions. A purely additive combination (e.g. Itti & Koch, 2000; Wolfe, 1994) can therefore be ruled out, and models assuming such a combination rule would need to be updated to account for the current results. It is possible that saliency is computed independently for each attribute, followed by a non-linear combination such as either a max rule (e.g. Zhaoping & May, 2007), or a compressive nonlinearity such as used here (e.g. our version of the Guided Search model).

## Implications for eye movements

One strategy to investigate saliency is to study eye movements during various tasks, indicative of information that both grabs the attention of the participant, and information that is used by the participant during the task (Itti & Koch, 2000; Krieger, Rentschler, Hauske, Schill & Zetsche, 2000; Mannan, Ruddock & Wooding, 1997; Parkhurst, Law & Niebur, 2002; Parkhurst & Niebur, 2003; Peters et al., 2005; Reinagel & Zador, 1999; Tatler, Baddeley & Gilchrist, 2005; Torralba, 2003; Torralba, Oliva, Castelano & Henderson, 2006). For example, saccade locations in natural scenes tend to contain more information, such as increased contrast (Krieger et al., 2000; Mannan, Ruddock & Wooding, 1996; Reinagl & Zador, 1999), changes in local luminance or contrast relative to the surround regions, and increased local structure such as variability in orientations at different spatial frequencies (as seen in corners and occlusions; Krieger et al., 2000) in contrast to single-orientation structure more common of natural image statistics when measured over comparable areas (Krieger et al., 2000; Krieger, Zetsche & Barth, 1997).

Saliency and saccades are usually linked by the assumption that highly salient image regions attract saccades, at least when top-down influences are small. Combining our

method to quantify saliency at all locations in an image with measurements of saccade endpoints within the same image would help refine this theory. For example, it is known that saccades are made to isolated saliency peaks. However, it is not known whether saccades are also made to regions surrounded by many saliency peaks even though the saccade end-point itself is not at a saliency peak. This strategy would be a good way to investigate several high-saliency points within a single saccade. This could occur, for example, if the map used to guide saccades to salient regions had a lower spatial resolution than the saliency map itself.

## Implications for contrast and duration

The method used in the present study is different from that used in previous experiments in several respects. Yet, when cross experiment comparisons are possible, the effects found here replicate those found in other experiments. The fact that saliency effects generalize well across experimental conditions has implications for our understanding of saliency. It seems that the same mechanisms compute saliency in a similar way over a relatively broad range of image contrasts, masking conditions, and presentation times. We cannot evaluate how our results might depend on spatial scaling (e.g. inter-item spacing, item size, and item bandwidth), as these factors were not systematically varied within the current experiment.

The present experiment measured saliency in relatively unchallenging exposure conditions (e.g. stimulus presentation was continuous, using large contrasts, task demands encouraged longer fixations between saccades). In contrast, other studies measured saliency effects using harder tasks, for example, by using short stimulus presentations (e.g. Nothdurft, 2000a, 2000b, 2002; van Zoest & Donk, 2004), using cues in difficult tasks (e.g. Nakayama & Mackeben, 1989; Nothdurft, 2002), measuring accuracy of first saccades (e.g. van Zoest & Donk, 2005, 2006; van Zoest, Donk & Theeuwes, 2004), and/or encouraging speeded responses (e.g. Koene & Zhaoping, 2007; Rutishauser & Koch, 2007; Treisman & Gelade, 1980; Treisman & Gormican, 1988; Treisman & Sato, 1990). In effect, whereas most tasks end either before or shortly after a saccade is made to the target or salient location, most of our measurements are made during a longer time period following saccades to salient locations. Our experiment was not designed for the purpose of measuring transient effects or comparing sustained and transient effects, and our instructions introduced a bias towards sustained saliency.

This distinction is relevant to the issue of whether saliency is transient or is inhibited. Indeed, competing theories postulate that saliency is transient (Nakayama & Mackeben, 1989; van Zoest & Donk, 2005, 2006) or that it is inhibited once the salient location has received attention (Itti et al., 1998; Itti & Koch, 2000; Wolfe & Gancarz, 1996). These theories agree that saliency should disappear at a location once a saccade is made to that location. Such

views however are inconsistent with the results from the present experiment, which show that saliency was sustained reliably in a static stimulus. Saliency was sustained to the extent that participants could locate a salient area, make a saccade to it, move the mouse pointer to the same region, and finally equalize luminance over a few seconds. Participants often would spend several seconds adjusting the luminance in an area while maintaining their fixation within that area. Therefore, purely transient accounts of saliency are at odds with our results. Even the transients introduced during a saccade to a new location would be gone by the time participants made the mouse pointer movements and adjustments to the luminance of the area. That is, even if present, transients triggered by saccades would not contribute significantly to their responses. This constitutes evidence for sustained saliency during fixation, at least for sustained stimulus presentations. We feel safe to conclude that purely transient accounts of saliency can be dismissed. Saliency either can be sustained, or it has sustained and transient components.

Some models include a mechanism that automatically inhibits saliency at a location once attention shifts to that location or a saccade has been generated to that location (Itti et al., 1998; Itti & Koch, 2000; Wolfe & Gancarz, 1996). However, these mechanisms are inconsistent with the current data. In the present experiment, participants reported looking at locations where they were making adjustments. That is, participants were making attentional shifts and saccades to the location they were adjusting, which should have inhibited saliency at that location, therefore eliminating any bias due to saliency from the participant's responses. That is, the above mechanisms predict that saliency is erased even before participants start adjusting luminance in a region, which is inconsistent with our data. It is thus clear that neither an attention shift nor a saccade to a location is sufficient to inhibit saliency at a location. However, these mechanisms can be modified to be consistent with the current data, i.e. either (1) inhibitory mechanisms suppress saliency but only once the attended location has been judged as irrelevant to the task, or (2) inhibitory mechanisms act on saccade-planning mechanisms rather than saliency mechanisms, therefore having no effect on perceived saliency.

The conclusion that purely transient accounts of saliency can be dismissed may seem difficult to reconcile with past research. However, caution must be taken when making inferences about saliency measurements based on task performance. There is no doubt that saliency influences performance. However, there are many ways that the spatiotemporal characteristics of *performance* and *saliency* might diverge; casting doubt on the frequent implicit assumption that saliency influences performance linearly.

The results reported by van Zoest and Donk (2006) may constitute an example of such divergence. They presented arrays of identical "non-target" oblique lines except for one vertical target line and one distractor line. Line orientations were chosen such that the distractor could ei-

ther be salient or not (i.e. the orientation difference between distractor and non-target lines was  $62.5^\circ$  or  $22.5^\circ$  respectively), and target-distractor similarity could either be high or low (i.e. the orientation difference between target and distractor was  $22.5^\circ$  or  $62.5^\circ$  respectively). The target line was always oriented  $45^\circ$  away from surrounding items (i.e. at an orientation difference midway between the distractor-surround orientation differences found in the high- and low-distractor-saliency conditions). Participants were asked to make a speeded saccade to the target. The observations were that early saccades ( $<250\text{ms}$ ) were made to salient items (i.e. incorrect saccades were mostly towards the highly-salient distractor) independently of target-distractor similarity, whereas late saccades ( $250\text{ms}$  to  $650\text{ms}$ ) were made towards items similar to the target (i.e. incorrect saccades were mostly towards the distractor if it was similar to the target) independently of distractor saliency. This was interpreted by the authors as meaning that "while the effect of distractor saliency was fast and transient, the effect of target-distractor similarity was slow and sustained" (p. 70). Thus they claim that saliency is transient. The alternative account we would suggest is that target and distractor saliencies increase rapidly, and fast saccades are frequently directed towards the most salient of the two items. However, if fast saccades are prevented, then saliency guides the identification mechanism to the target and distractor items, such that the target can be correctly identified. That is, by our account, the performance of the identification mechanism relies on saliency to single out these two items for further analysis, but which of these two items is the most salient is irrelevant to the identification mechanism. Saccade accuracy in their experiment for those slow saccades is only dependent on the performance of the identification mechanism, as both the target and the distractor were salient.

Note that by our account of van Zoest and Donk's (2006) results, saliency improves performance for both fast and slow saccades. Indeed, reducing target saliency by making the target item more similar to non-targets can only decrease performance, as participants will make more saccades either to non-targets (these saccades were removed from their analyses) or to the distractor. Without sustained saliency at the target location, it is unclear how participants could even make correct saccades at longer durations.

As a second example, Nakayama and Mackeben (1989) measured the effect of spatial cueing on performance in a task where participants had to identify if there was an oddball item, and if so, to report its color. The spatial cueing they used was a frame surrounding the item of interest. They found that cueing did not improve task performance when the oddball item had a unique defining feature (i.e. orientation). They argue that the target item acted as its own cue, and that cueing did not improve performance even though it might have increased the target's saliency.

Nakayama and Mackeben also measured performance in a similar task where the oddball was defined as a conjunction of features, and found a transient improvement of

performance when the oddball was presented shortly after cue onset. They argued that the cue triggered transient saliency. However, the short-lived increase in performance following cueing can be equally well explained as a combination of two effects (see Poirier & Frost, 2005): (1) sustained and spatially-broad “integration” mechanisms operating on objects, and (2) delayed inhibitory or “segregation” mechanisms operating between objects. The key to understanding this is that in all of Nakayama and Mackeben’s experiments, the cue was a frame surrounding the target item location, and the target was presented after the frame’s onset. By our interpretation, the frame’s saliency increased monotonically with time, with some of that saliency leaking from the frame to the target, thus improving performance. However, this benefit would be short-lived either because the saliency effect became specific to the frame over time, or because the salient frame masked the target after some delay. This theory explains the observed performance rise and fall shortly after cue presentation even though the cue’s saliency increases monotonically.

Nakayama and Mackeben also tried to lengthen the duration of the transient performance increase. Specifically, they made the cue flicker, thus argued that the cue remained salient which should also maintain the performance increase due to saliency. However, they found that the performance increase remained transient. That is, trying to maintain the cue’s saliency by making it flicker failed to change the pattern of results. By our account, this manipulation failed to change the pattern of results because the cue’s masking effect remained.

Thus, in experiments conducted by both van Zoest and Donk (2006) and by Nakayama and Mackeben (1989), it is difficult to make solid conclusions regarding whether saliency is transient or sustained on the basis of performance measures because of several interfering factors which affect such measures. In particular, their data does not rule out the presence of sustained saliency. In contrast, our findings do rule out purely transient accounts of saliency, as explained above. What remains to be determined is whether saliency is purely sustained, or has both sustained and transient components.

### Was saliency really measured?

Did we measure saliency rather than something else that is correlated with saliency? Participants adjusted luminance rather than saliency, and search performance was not measured. We therefore need to assess whether the visual spread task itself suffers from extraneous variables, and more fundamentally, if it did measure saliency rather than luminance effects or some other perceptual effect(s).

The visual spread task is less influenced by extraneous factors by virtue of having simpler task demands than most other tasks. For example, Nothdurft’s measurement of saliency (1993b, 2000a) has been challenged (Huang & Pashler, 2005; Koene & Zhaoping, 2007) on the grounds that saliency is an abstract concept, for which participants have

difficulty making reliable judgments (Koene & Zhaoping, 2007). However, in the visual spread task, participants were asked to make luminance judgments rather than saliency judgments. Luminance judgments are more accurate and intuitive than saliency judgments. As a further example, participants did not search for a target in the visual spread task. That is, there were no explicit incentives for applying top-down biases to saliency (e.g. Bacon & Egeth, 1997; Lamy, Leber & Egeth, 2004; Leber & Egeth, 2006; Sobel & Cave, 2002). Thus, our measure is less likely to be influenced by strategies than performance-based saliency measurements. In short, *if saliency was indeed measured here*, it was less likely to be influenced by extraneous variables than other known methods of measuring saliency.

In the visual spread task, participants adjusted the luminance of lines on a black background, which also influenced local contrast. There is no doubt that luminance and contrast can influence visibility, saliency, and ultimately performance. But the converse is not necessarily true: it is possible that saliency did not influence perceived luminance, thus that the adjustments made by participants were not influenced by saliency. Further experiments will be necessary to validate the visual spread task as a measure of saliency.

## Conclusions

The efficiency of the visual spread method for data collection is noteworthy. This efficiency is gained by combining the advantages of methods of adjustment to the ability to make adjustments at any location in the image. Essentially, participants were free to concentrate their responses at locations that deviated most from apparent equiluminance, thus quickly reducing apparent luminance variations throughout the stimulus. This method quickly converged on luminance settings that were similar across participants, and once analyzed showed similar effects to those reported in the literature. Moreover, the data collected was correlated with predictions from a model, which strengthens both lines of research. Our method was also sensitive enough to detect and characterize nonlinearities in the function relating saliency to the similarity of nearby items, as well as interactions between dimensions, and the enhancement of both collinear and end-stopping items.

The possible applications for such a reliable and rapid data-collection method are widespread. We are currently adapting the method for collecting data on other stimulus classes, including text (reading) and natural images (for a demonstration, see Figure 1).

## Acknowledgments

This research was supported by a CIHR grant awarded to Martin Arguin, Frédéric Gosselin, and Dan Bub. Portions of this paper were presented at VSS 2007 (Sarasota, Florida), Tennen 2007 (Montréal, Canada), and Cernec

2007 (Montréal, Canada). We thank Nicolas Dupuis-Roy, Zakia Hammal, the editor, and 2 reviewers for comments on the manuscript.

Commercial relationships: none.

Corresponding author: Frédéric J.A.M. Poirier.

Email: [frederic.poirier@umontreal.ca](mailto:frederic.poirier@umontreal.ca).

Address: C.P. 6128, succursale Centre-Ville, Montréal, Québec, Canada, H3C 3J7.

## References

- Abrams, R. A., & Christ, S. E. (2005). The onset of receding motion captures attention: Comment on Franconeri and Simons (2003). *Perception & Psychophysics*, *67* (2), 219-223. [PubMed]
- Assad, J.A. (2003). Neural coding of behavioral relevance in parietal cortex. *Current Opinion in Neurobiology*, *13*, 194-197. [PubMed]
- Bacon, W.J., & Egeth, H.E. (1997). Goal-directed guidance of attention: evidence from conjunctive visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *23*(4), 948-961. [PubMed]
- Blasdel, G. (1992). Orientation selectivity, preference, and continuity in monkey striate cortex. *Journal of Neuroscience*, *12* (8), 3139-3161. [PubMed]
- Blaser, E., Pylyshyn, Z.W., & Holcombe, A.O. (2000). Tracking an object through feature-space. *Nature*, *408*(6809), 196-199. [PubMed]
- Braun, J. (1994). Visual search among items of different salience: Removal of visual attention mimics a lesion in extrastriate area V4. *Journal of Neuroscience*, *14*, 554-567. [PubMed]
- Brown, W. (1910). Some experimental results in the correlation of mental abilities. *British Journal of Psychology*, *3*, 296-322.
- Carmi, R., & Itti, L. (2006). Causal saliency effects during natural vision. *Proceedings of the 2006 symposium on Eye tracking research & applications*, 11-18.
- Colby, C.L., & Goldberg, M.E. (1999). Space and attention in parietal cortex. *Annual Review of Neuroscience*, *22*, 319-349. [PubMed]
- Das, A., & Gilbert, C. (1999). Topography of contextual modulations mediated by short-range interactions in primary visual cortex. *Nature*, *399* (6737), 655-661. [PubMed]
- Einhäuser, W., & König, P. (2003). Does luminance-contrast contribute to a saliency map for overt visual attention? *European Journal of Neuroscience*, *17*, 1089-1097. [PubMed]
- Folk, C. L., Remington, R. W., & Johnston, J. C. (1992). Involuntary covert orienting is contingent on attentional control settings. *Journal of Experimental Psychology: Human Perception and Performance* *18*, 1030-1044. [PubMed]
- Foster, D.H., & Ward, P.A. (1991). Asymmetries in oriented-line detection indicate two orthogonal filters in early vision. *Proceedings of the Royal Society of London, B*, *243*, 75-81. [PubMed]
- Franconeri, S.L., Hollingworth, A., & Simons, D.J. (2005). Do new objects capture attention? *Psychological Science*, *16*, 275-281. [PubMed]
- Gilbert, C.D. (1992). Horizontal integration and cortical dynamics. *Neuron*, *9*, 1-13. [PubMed]
- Gottlieb, J.P., Kusunoki, M., & Goldberg, M.E. (1998). The representation of visual salience in monkey parietal cortex. *Nature*, *391*, 481-484. [PubMed]
- Hillstrom, A.P., & Yantis, S. (1994). Visual motion and attentional capture. *Perception and Psychophysics*, *55*, 399-411. [PubMed]
- Horwitz, G.D., & Newsome, W.T. (1999). Separate signals for target selection and movement specification in the superior colliculus. *Science*, *284*(5417), 1158-1161. [PubMed]
- Huang, L., & Pashler, H. (2005). Quantifying object salience by equating distractor effects. *Vision Research*, *45*(14), 1909-1920. [PubMed]
- Hubel, D.H., & Wiesel, T.N. (1965). Receptive fields and functional architecture in two non-striate visual areas (18 and 19) of the cat. *Journal of Neurophysiology*, *28*, 229-289. [PubMed]
- Itti, L. (2006). Quantitative modeling of perceptual salience at human eye position. *Visual Cognition*, *14* (4-8), 959-984.
- Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *20* (11), 1254-1259.
- Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, *40* (10-12), 1489-1506. [PubMed]
- Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience*, *2* (3), 194-203. [PubMed]
- Jingling, L., & Zhaoping, L. (2008). Change detection is easier at texture border bars when they are parallel to the border: Evidence for V1 mechanisms of bottom-up salience. *Perception*, *37*, 197-206. [PubMed]
- Jonides, J., & Yantis, S. (1988). Uniqueness of Abrupt Visual Onset in Capturing Attention. *Perception and Psychophysics*, *43*, 346-354. [PubMed]

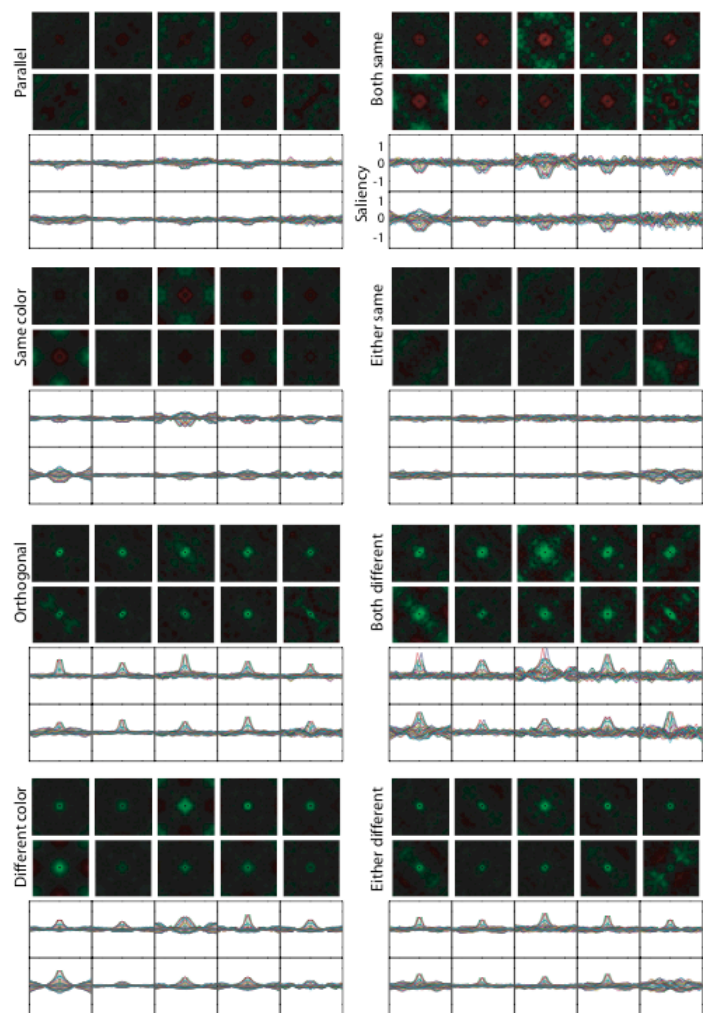
- Kapadia, M.K., Ito, M., Gilbert, C.D., & Westheimer, G. (1995). Improvement in visual sensitivity by changes in local context: Parallel studies in human observers and in V1 of alert monkeys. *Neuron*, *15*, 843-856. [PubMed]
- Kim, M. S., & Cave, K. R. (1999). Top-down and bottom-up attentional control: On the nature of interference from a salient distractor. *Perception and Psychophysics*, *61*, 1009-1023. [PubMed]
- Knierim, J.J., & Van Essen, D.C. (1992). Neuronal responses to static texture patterns in area V1 of the alert macaque monkey. *Journal of Neurophysiology*, *67*(4), 961-980. [PubMed]
- Koene, A. R., & Zhaoping, L. (2007). Feature-specific interactions in salience from combined feature contrasts: Evidence for a bottom-up saliency map in V1. *Journal of Vision*, *7*(7):6, 1-14, <http://journalofvision.org/7/7/6/>, doi:10.1167/7.7.6. [PubMed] [Article]
- Krieger, G., Rentschler, I., Hauske, G., Schill, K., & Zetzsche, C. (2000). Object and scene analysis by saccadic eye-movements: an investigation with higher-order statistics. *Spatial Vision*, *13*, 201-214. [PubMed]
- Krieger, G., Zetzsche, C., & Barth, E. (1997). Higher-order statistics of natural images and their exploitation by operators selective to intrinsic dimensionality. In *Proceedings of the 1997 IEEE Signal Processing Workshop on Higher-Order Statistics*, 147-151. [PubMed]
- Kustov, A.A., & Robinson, D.L. (1996). Shared neural control of attentional shifts and eye movements. *Nature*, *384* (6604), 74-77. [PubMed]
- Lamy, D., Leber, A., & Egeth, H.E. (2004). Effects of task relevance and stimulus-driven salience in feature-search mode. *Journal of Experimental Psychology: Human Perception and Performance*, *30*(6), 1019-1031. [PubMed]
- Leber, A.B., & Egeth, H.E. (2006). It's under control: top-down search strategies can override attentional capture. *Psychonomic Bulletin & Review*, *13*(1), 132-138. [PubMed]
- Lee, D.K., Itti, L., Koch, C., & Braun, J. (1999). Attention activates winner-take-all competition among visual filters. *Nature Neuroscience*, *2*, 375-381. [PubMed]
- Li, Z. (1998). A neural model of contour integration in the primary visual cortex. *Neural Computation*, *10*, 903-940. [PubMed]
- Li, Z. (1999). Contextual influences in V1 as a basis for pop out and asymmetry in visual search. *Proceedings of the National Academy of Sciences, USA*, *96*, 10530-10535. [PubMed]
- Li, Z. (2002). A saliency map in primary visual cortex. *Trends in Cognitive Sciences*, *6* (1), 9-16. [PubMed]
- Mannan, S.K., Ruddock, K. & Wooding, D. (1996). The relationship between the locations of spatial features and those of fixations made during visual examination of briefly presented images. *Spatial Vision*, *10*, 165-188. [PubMed]
- Mannan, S. K., Ruddock, K. H., & Wooding, D. S. (1997). Fixation patterns made during brief examination of two-dimensional images. *Perception*, *26*, 1059-1072. [PubMed]
- McAdams, C.J., & Maunsell, J.H.R. (2000). Attention to both space and feature modulates neuronal responses in macaque area V4. *Journal of Neurophysiology*, *83*, 1751-1755. [PubMed]
- McPeck, R.M., & Keller, E.L. (2002). Saccade target selection in the superior colliculus during a visual search task. *Journal of Neurophysiology*, *88*, 2019-2034. [PubMed]
- Nakayama, K., & Mackeben, M. (1989). Sustained and transient components of focal visual attention. *Vision Research*, *29*(11), 1631-47. [PubMed]
- Nelson, J.I., & Frost, B.J. (1985). Intracortical facilitation among co-oriented, co-axially aligned simple cells in cat striate cortex. *Experimental Brain Research*, *61*, 54-61. [PubMed]
- Nothdurft, H.-C. (1992). Feature analysis and the role of similarity in preattentive vision. *Perception and Psychophysics*, *52*, 355-375. [PubMed]
- Nothdurft, H.-C. (1993a). Saliency effects across dimensions in visual search. *Vision Research*, *33*, 839-844. [PubMed]
- Nothdurft, H.-C. (1993b). The conspicuousness of orientation and motion contrast. *Spatial Vision*, *7*, 341-363. [PubMed]
- Nothdurft, H.-C. (2000a). Saliency from feature contrast: additivity across dimensions. *Vision Research*, *40*, 1183-1201. [PubMed]
- Nothdurft, H.-C. (2000b). Saliency from feature contrast: variations with texture density. *Vision Research*, *40*, 3181-3200. [PubMed]
- Nothdurft, H.-C. (2002). Attention shifts to salient targets. *Vision Research*, *42*, 1287-1306. [PubMed]
- Nothdurft, H.-C. (2006). Saliency and target selection in visual search. *Visual Cognition*, *14* (4-8), 514-542.
- O'Craven, K.M., Downing, P.E., & Kanwisher, N. (1999). fMRI evidence for objects as the units of attentional selection. *Nature*, *401*, 584-587. [PubMed]
- Olmos, A., & Kingdom, F.A.A. (2004). McGill Calibrated Colour Image Database, <http://tabby.vision.mcgill.ca>.
- Parkhurst, D., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision Research* *42*, 107-123. [PubMed]

- Parkhurst, D. J., & Niebur, E. (2003). Scene content selected by active vision. *Spatial Vision*, *16*, 125-154. [PubMed]
- Peters, R. J., Iyer, A., Itti, L., & Koch, C. (2005). Components of bottom-up gaze allocation in natural images. *Vision Research*, *45*, 2397-2416. [PubMed]
- Pettet, M., & Gilbert, C. (1992). Dynamic changes in receptive-field size in cat primary visual-cortex. *Proceedings of The National Academy of Sciences of the United States of America*, *89* (17), 8366-8370. [PubMed]
- Poirier, F.J.A.M., & Frost, B.J. (2005). Global orientation aftereffect in multi-attribute displays: implications for the binding problem. *Vision Research*, *45* (4), 497-506. [PubMed]
- Poirier, F.J.A.M., & Gurnsey, R. (1998). The effects of eccentricity and spatial frequency on the orientation discrimination asymmetry. *Spatial Vision*, *11*, 349-366. [PubMed]
- Polat, U., & Bonneh, Y. (2000). Collinear interactions and contour integration. *Spatial Vision*, *13* (4), 393-401. [PubMed]
- Polat, U., Mizobe, K., Pettet, M., Kasamatsu, T. & Norcia A. M. (1998). Collinear stimuli regulate visual responses depending on cell's contrast threshold. *Nature*, *391*, 580-584. [PubMed]
- Polat, U. & Norcia A. M. (1996). Neurophysiological evidence for contrast dependent long range facilitation and suppression in the human visual cortex. *Vision Research*, *36*, 2099-2109. [PubMed]
- Polat, U. & Norcia, A. M. (1998). Elongated physiological summation pools in the human visual cortex. *Vision Research*, *38*, 3735-3741. [PubMed]
- Polat, U. & Sagi, D. (1993). Lateral interactions between spatial channels: suppression and facilitation revealed by lateral masking experiments. *Vision Research*, *33*, 993-999. [PubMed]
- Polat, U. & Sagi, D. (1994a). The architecture of perceptual spatial interactions. *Vision Research*, *34*, 73-78. [PubMed]
- Polat, U. & Sagi, D. (1994b). Spatial interactions in human vision: From near to far via experience-dependent cascades of connections, *Proceedings of the National Academy of Sciences, USA*, *91*, 1206-1209. [PubMed]
- Posner, M.I., & Peterson, S.E. (1990). The attention system of the human brain. *Annual Review of Neuroscience*, *13*, 25-42. [PubMed]
- Reinagl, P. & Zador, A. M. (1999). Natural scene statistics at the centre of gaze. *Network: Computations in Neural Systems*, *10*, 341- 350. [PubMed]
- Reynolds, J.H., Alborzian, S., & Stoner, G.R. (2003). Exogenously cued attention triggers competitive selection of surfaces. *Vision Research*, *43*, 59-66. [PubMed]
- Robinson, D.L., & Peterson, S.E. (1992). The pulvinar and visual salience. *Trends in Neuroscience*, *15*(4), 127-132. [PubMed]
- Rockland, K.S., Andresen, J., Cowie, R.J., & Robinson, D.L. (1999). Single axon analysis of pulvinocortical connections to several visual areas in the macaque. *The Journal of Comparative Neurology*, *406*(2), 221-250. [PubMed]
- Rockland, K.S., & Lund, J.L. (1983). Intrinsic laminar lattice connections in primate visual cortex. *Journal of Comparative Neurology*, *216*, 303-318. [PubMed]
- Rutishauser, U., & Koch, C. (2007). Probabilistic modeling of eye movement data during conjunction search via feature-based attention. *Journal of Vision*, *7*(6):5, 1-20, <http://journalofvision.org/7/6/5/>, doi:10.1167/7.6.5. [PubMed] [Article]
- Saenz, M., Buracas, G.T., & Boynton, G.M. (2002). Global effects of feature-based attention in human visual cortex. *Nature Neurosciences*, *5*, 631-632. [PubMed]
- Sagi, D., & Julesz, B. (1985). "Where" and "what" in vision. *Science*, *228*(4704), 1217-1219. [PubMed]
- Schall, J.D. (2002). The neural selection and control of saccades by the frontal eye field. *Philosophical Transactions of the Royal Society of London B, Biological Sciences*, *357*, 1073-1082. [PubMed]
- Sobel, K. V., & Cave, K. R. (2002). Roles of salience and strategy in conjunction search. *Journal of Experimental Psychology: Human Perception and Performance*, *28*, 1055-1070. [PubMed]
- Spearman, C. (1910). Correlation calculated with faulty data. *British Journal of Psychology*, *3*, 271-295.
- Stettler, D., Das, A., Bennett, J., & Gilbert, C. (2002). Lateral connectivity and contextual interactions in macaque primary visual cortex. *Neuron*, *36*(4), 739-750. [PubMed]
- 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]
- Theeuwes, J. (1991). Exogenous and endogenous control of attention: The effect of visual onsets and offsets. *Perception & Psychophysics*, *49*, 83-90. [PubMed]
- Theeuwes, J. (1994). Stimulus-driven capture and attentional set-selective search for color and visual abrupt onsets. *Journal of Experimental Psychology: Human Perception and Performance* *20*, 799-806. [PubMed]
- Thompson, K.G., Bichot, N.P., & Schall, J.D. (1997). Dissociation of visual discrimination from saccade programming in macaque frontal eye field. *Journal of Neurophysiology*, *77*, 1046-1050. [PubMed]
- Torralba, A. (2003). Modeling global scene factors in attention. *Journal of the Optical Society of America A. Optics, Image Science, and Vision* *20*, 1407-1418. [PubMed]



- Torralba, A., Oliva, A., Castelano, 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 (4), 766-786. [PubMed]
- Treisman, A., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, 12, 97-136. [PubMed]
- Treisman, A., & Gormican, S. (1988). Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review*, 95 (1), 15-48. [PubMed]
- Treisman, A., & Sato, S. (1990). Conjunction search revisited. *Journal of Experimental Perception and Performance*, 16, 459-478. [PubMed]
- Treisman, A., & Souther, J. (1985). Search asymmetry: a diagnostic for preattentive processing of separable features. *Journal of Experimental Psychology: General*, 114, 285-310. [PubMed]
- Treue, S. (2001). Neural correlates of attention in primate visual cortex. *Trends in Neurosciences*, 24 (5), 295-300. [PubMed]
- Treue, S. (2003). Visual attention: the where, what, how and why of saliency. *Current Opinion in Neurobiology*, 13 (4), 428-432. [PubMed]
- Treue, S., & Martinez-Trujillo, J.C. (1999). Feature-based attention influences motion processing gain in macaque visual cortex. *Nature*, 399, 575-579. [PubMed]
- Tsotsos, J.K., Culhane, S.M., Wai, W.Y.K., Lai, Y., Davis, N., & Nuflo, F. (1995). Modeling visual attention via selective tuning. *Artificial Intelligence*, 78 (1-2), 507-545.
- van Zoest, W., & Donk, M. (2004). Bottom-up and top-down control in visual search. *Perception*, 33 (8), 927-937. [PubMed]
- van Zoest, W. & Donk, M. (2005). The effects of salience on saccadic target selection. *Visual Cognition*, 2 (2), 353-375.
- van Zoest, W., & Donk, M. (2006). Saccadic target selection as a function of time. *Spatial Vision*. 19 (1), 61-67. [PubMed]
- van Zoest, W., Donk, M. & Theeuwes, J. (2004). The role of stimulus-driven and goal-driven control in saccadic visual selection. *Journal of Experimental Psychology: Human Perception and Performance*, 30 (4), 746-759. [PubMed]
- Wachtler, T., Sejnowski, T.J., & Albright, T.D. (2003). Representation of color stimuli in awake macaque primary visual cortex. *Neuron*, 37(4), 681-691. [PubMed]
- Wolfe, J.M. (1994). Guided Search 2.0: A revised model of visual search. *Psychonomic Bulletin & Review*, 1 (2), 202-238.
- Wolfe, J. M., & Gancarz, G. (1996). Guided Search 3.0: A model of visual search catches up with Jay Enoch 40 years later. In V. Lakshminarayanan (Ed.), *Basic and Clinical Applications of Vision Science* (pp. 189-192). Dordrecht, Netherlands: Kluwer Academic.
- Yantis, S., and Egeth, H. E. (1999). On the distinction between visual salience and stimulus-driven attentional capture. *Journal of Experimental Psychology: Human Perception and Performance*, 25, 661-676. [PubMed]
- Zhaoping, L., & May, K.A. (2007). Psychophysical tests of the hypothesis of a bottom-up saliency map in primary visual cortex. *PLoS Computational Biology*, 3 (4), e62. [PubMed]

## Sup. figure caption



**Supplementary Figure 1.** Same as Figure 6, except that analyses are shown for each of the 10 participants, and predictions are not shown. Despite some differences in amplitudes and signal-to-noise ratios, participants are remarkably consistent with each other. See Figure 6 caption for details.

