

# What's color got to do with it? The influence of color on visual attention in different categories

**Hans-Peter Frey**

Institute of Cognitive Science, University of Osnabrück,  
Osnabrück, Germany



**Christian Honey**

Institute of Cognitive Science, University of Osnabrück,  
Osnabrück, Germany



**Peter König**

Institute of Cognitive Science, University of Osnabrück,  
Osnabrück, Germany



Certain locations attract human gaze in natural visual scenes. Are there measurable features, which distinguish these locations from others? While there has been extensive research on luminance-defined features, only few studies have examined the influence of color on overt attention. In this study, we addressed this question by presenting color-calibrated stimuli and analyzing color features that are known to be relevant for the responses of LGN neurons. We recorded eye movements of 15 human subjects freely viewing colored and grayscale images of seven different categories. All images were also analyzed by the saliency map model (L. Itti, C. Koch, & E. Niebur, 1998). We find that human fixation locations differ between colored and grayscale versions of the same image much more than predicted by the saliency map. Examining the influence of various color features on overt attention, we find two extreme categories: while in rainforest images all color features are salient, none is salient in fractals. In all other categories, color features are selectively salient. This shows that the influence of color on overt attention depends on the type of image. Also, it is crucial to analyze neurophysiologically relevant color features for quantifying the influence of color on attention.

Keywords: attention, eye movements, color vision, image features, natural stimuli

Citation: Frey, H.-P., Honey, C., & König, P. (2008). What's color got to do with it? The influence of color on visual attention in different categories. *Journal of Vision*, 8(14):6, 1–17, <http://journalofvision.org/8/14/6/>, doi:10.1167/8.14.6.

## Introduction

The visual environment normally encountered by humans is complex. It is not possible for the human brain to simultaneously process all incoming visual information, so it sequentially targets discrete parts of the environment for closer analysis. Attention thus allows us to dissect complex visual input into manageable portions.

Typically, a distinction is made between covert and overt visual attention, based on the role of eye movements. The former refers to a shift of attention without a corresponding shift of gaze, and a first description of this phenomenon dates back to von Helmholtz (1867). The latter is related to eye movements and involves directing the gaze to interesting—or salient—locations. However, it has been shown that eye movements and attention are correlated in human subjects (Hoffman & Subramaniam, 1995; Maioli, Benaglio, Siri, Sosta, & Cappa, 2001). Furthermore, animal experiments have found cells in superior colliculus that are active both during saccade preparation and covert shifts of attention (Ignashchenkova, Dicke, Haarmeier, & Thier, 2004; Kustov & Robinson, 1996), indicating that there is also a common neuronal

substrate. The analysis of eye movements therefore provides an objective measure of attentional processes.

Before an eye movement occurs, the pre-attentive conspicuity, or saliency, of regions of the visual scene must be calculated. Several electrophysiological studies have found neural correlates of this saliency calculation in brain regions like the pulvinar (Posner & Petersen, 1990), the frontal eye field (Thompson, Bichot, & Schall, 1997), superior colliculus (Horwitz & Newsome, 1999), and the lateral intraparietal area (Gottlieb, Kusunoki, & Goldberg, 1998). A recent study by Mazer and Gallant (2003) of macaque monkeys viewing grayscale natural scenes indicated that the ventral areas V4 and IT are involved in the computation of saliency. Macaque V4 (Zeki, 1983) and IT (Komatsu, Ideura, Kaji, & Yamane, 1992) are also associated with the processing of color. This guides our interest toward the relation between color and attention.

There are three different kinds of color photoreceptors (cones) in the normal human retina, which respond preferentially to different wavelengths of visible light: short (S, whose absorption spectrum has a maximum at 440 nm), middle (M, most sensitive to wavelengths around 535 nm), and long (L, 565 nm). The color of an object can only be computed unambiguously if the

magnitudes of the outputs of all three cone types are compared. This processing is carried out by the horizontal and ganglion cells of the retina (Gegenfurtner & Kiper, 2003). Further processing takes place by means of two color opponent mechanisms in the parvocellular layers of the lateral geniculate nucleus (LGN) and an achromatic opponent mechanism in the magnocellular layers of the LGN (Derrington, Krauskopf, & Lennie, 1984). Receptive fields of opponent cells are composed of a center and a surround, which are spatially antagonistic (Gegenfurtner & Kiper, 2003). Saturation modulates the firing rates of the color-opponent cells in the LGN. The achromatic opponent mechanism refers to cells that are excited or inhibited by the presence or absence of light in the center or the surround of their receptive field. These results are taken to be valid for extrapolation to the human visual system.

Humans and few other primates are the only trichromatic mammals. They have a subsystem for comparing the outputs of middle (M)- and long-wavelength-sensitive (L) cones (Nathans, 1999), which means that they can discriminate well between red and green. Trichromacy evolved only about 30–40 million years ago in the Old World primate lineage. Hypotheses for the evolution and maintenance of trichromacy emphasize its role in the ability to forage for edible fruits (Sumner & Mollon, 2000) or young leaves (Dominy & Lucas, 2001; Sumner & Mollon, 2000). These studies showed that the visual system of trichromatic primates is optimally tuned to discriminate edible fruits and young leaves from their natural background. Most of these studies were conducted using spectral measurements from the Kibale Rainforest in Uganda, so in order to analyze the salience of color features in a setting in which trichromatic color vision is advantageous, we use calibrated color images acquired in the same rainforest environment. We expect that for images in this category, the red–green color subsystem will influence fixation behavior more than in the other image categories used in this study.

Where we direct our gaze depends on expectations, experience, and the experimental task (top-down aspects), as well as the properties or intrinsic features of the stimulus like brightness, color, or movement (bottom-up aspects). The influences of top-down processing have been examined at least since Buswell's (1935) study. In the case of bottom-up attentional processes, two distinct but complementary approaches are generally applied: biologically inspired modeling and statistical approaches.

Based on neurophysiological and psychophysical findings, Koch and Ullman (1985) proposed the first version of a biologically plausible model for bottom-up overt attention: the saliency map. This model has undergone several different implementations; however, its basic scheme remains unchanged (for a review, see Itti & Koch, 2001). The stimulus is analyzed in various feature channels like luminance, color, orientation, or motion. Color processing is implemented in two channels, which

mimic color-opponent pathways in trichromatic primates. In each feature channel, local differences are computed, combined across several spatial scales and normalized in a nonlinear way. These “conspicuity maps” (Itti, Koch, & Niebur, 1998) are then summed up to yield the saliency map. Locations of high activity in the map are assumed to be salient, i.e., highly likely to be attended. The success of the model can be determined by examining its performance in predicting fixations of human observers. In the case of still images, namely grayscale outdoor scenes (Peters, Iyer, Itti, & Koch, 2005) and colored fractals, home interiors, landscapes, and outdoor scenes (Parkhurst, Law, & Niebur, 2002), neurobiologically plausible models were able to predict fixations to a certain extent. The ability of such models to discriminate between fixated and control image regions has also been found to be higher than chance (Kienzle, Wichmann, Schölkopf, & Franz, 2007). The saliency map approach has also been applied to movie clips and was found to predict fixation targets well above chance (Carmi & Itti, 2006; Le Meur, Le Callet, & Barba, 2007). These results suggest that neurobiologically inspired models can discriminate between fixated and non-fixated regions.

The second and somewhat younger approach is based on the statistical structure of the stimulus. The visual system selectively samples the natural environment at a rate of about 3 fixations per second. In grayscale images, it has been found that the image statistics at fixated regions differ from those at non-fixated locations, for example in luminance contrast (Reinagel & Zador, 1999), edge density (Mannan, Ruddock, & Wooding, 1996), and 2nd order luminance contrast (“texture contrast,” Parkhurst & Niebur, 2004). In colored images, it was shown that chromaticity is a predictive feature (Tatler, Baddeley, & Gilchrist, 2005), but its salience differs between image categories (Parkhurst et al., 2002). These studies show that we can find local operators that are able to predict, to a certain extent, where human subjects fixate. It should be noted, however, that these studies only deliver correlative analyses—eye-tracking studies using modified stimuli have shown that luminance contrast in the range of natural variations does not causally attract overt attention (Einhäuser & König, 2003). Therefore, we have to pay special attention to correlative effects when analyzing the salience of features.

In our study, we employ both of the bottom-up approaches mentioned above in order to determine the influence of color information on overt visual attention. We measure eye movements of human subjects while they look at images of seven categories (*Face, Flower & Animal, Forest, Fractal, Landscape, Man-Made, and Rainforest*) in two different conditions (colored and grayscale). These stimuli are defined in a neurophysiologically plausible color space, which models the responses of LGN cells (Derrington et al., 1984). Using only a minimal instruction, we try to reduce task-related top-down influences. We then apply the saliency map

model to exactly the same stimuli. Comparing fixations and model-predicted fixation locations between the two conditions reveals a general influence of color on selection of fixation locations since all other image features remain the same. If we find this general influence, we have to further break it down into its constituents. We then can analyze whether color information draws all subjects' gaze to similar locations, i.e., whether it causally attracts attention. The last step is to analyze which color features are salient. Up to now, there has been no systematic study of the influence of color features on overt attention. In the published studies that employ different categories of stimuli, color information is reduced to a single feature. However, color information can be described using several features like saturation or color contrast in the RG and BY color channels. In a recent study, we found that these different color features are not salient in naturally colored but are in color-modified versions of middle European landscapes (Frey, König, & Einhäuser, 2007). It is however probable that different color features are selectively salient in different types of environments. We thus divide natural scenes into seven categories according to semantic aspects and analyze the salience of neurophysiologically plausible color features within these different categories.

## Methods

### Subjects

Fifteen undergraduate students from the University of Osnabrück participated in the experiment. All subjects had normal or corrected-to-normal visual acuity. Each subject was tested for normal color vision using the Ishihara test for color deficiency (Kanohara Trading, Tokyo, Japan). They had not seen the stimuli before and were naïve to our specific research questions. All subjects gave written informed consent to participate in the experiment. The experiment conformed to the Declaration of Helsinki.

### Color representation

Color images can be represented using either neurophysiologically or psychophysically defined color-spaces. We believe that neurophysiologically defined spaces are more appropriate for this purpose (Frey et al., 2007). One such color space is DKL space (Derrington et al., 1984), which is based on the relative excitations of the three cone types (L, M, and S) in the retina of non-human primates. Three orthogonal axes constitute this color space: (1) “Constant blue” is given by the difference between L and M cone excitations (L–M). For the sake of simplicity, we will refer to this axis as the red–green (RG) axis. (2)

“Tritanopic confusion” is defined by  $(L + M) - S$ . We will refer to this axis as the blue–yellow (BY) axis. (3) “Luminance” is defined by  $(L + M)$ .

The azimuth in the plane of the two color axes defines a color's hue ( $0^\circ$  at  $RG > 0$ ,  $BY = 0$ ). The projection of a pixel in DKL space onto this isoluminant color plane (luminance = 0) preserves the chromatic properties of the pixel and we refer to the result of this projection as the chromatic content of a pixel.

### Stimuli

We used 191 images from 7 different scene categories: *Face* (26), *Flower & Animal* (30), *Forest* (30), *Fractal* (25), *Landscape* (19), *Man-Made* (32), and *Rainforest* (29).

*Face* stimuli included frontal, close-up shots of faces, taken indoors with a high-resolution digital camera (Sony DSC-V1 Cyber-Shot, Tokyo, Japan) under artificial lighting conditions (Acik et al, submitted). *Fractal* stimuli consisted of pictures taken from a World Wide Web database (<http://www.cnspace.net/html/fractals.html>) of software-generated fractals. Images from the Kibale Forest image data set (Troscianko et al., 2003) were used for the rainforest category, and stimuli for all remaining categories were sourced from the McGill Calibrated Color Images Database (Olmos & Kingdom, 2004). One example image from each category is depicted in *Figure 1* (panels A–G).

The images were down-sampled to a resolution of  $1024 \times 768$  pixels using bicubic interpolation. Each image was presented in two conditions: once colored and once in grayscale. Luminance in DKL space is given by the value along the luminance axis, and to generate the grayscale images from the original colored images, we transformed the DKL luminance information to RGB. An example grayscale image from the rainforest category is shown in *Figure 1* (panel H).

For stimulus presentation, we used a 21-in. CRT monitor (SyncMaster 1100 DF, Samsung Electronics, Suwon, South Korea; CIE coordinates of the phosphors: red 0.628/0.328, green 0.28/0.598, blue 0.146/0.06) at 100-Hz vertical refresh rate. Gamma of the presentation monitor was corrected in order to achieve a linear mapping of DKL values to monitor output. Subjects were seated 80 cm from the monitor surface, which yielded approximately  $28 \times 21$  degrees of visual angle for our stimuli.

### Eye tracking

For recording eye movements, we used the EyeLink II system (SR Research, Ontario, Canada). This head-mounted device uses two video cameras to monitor the

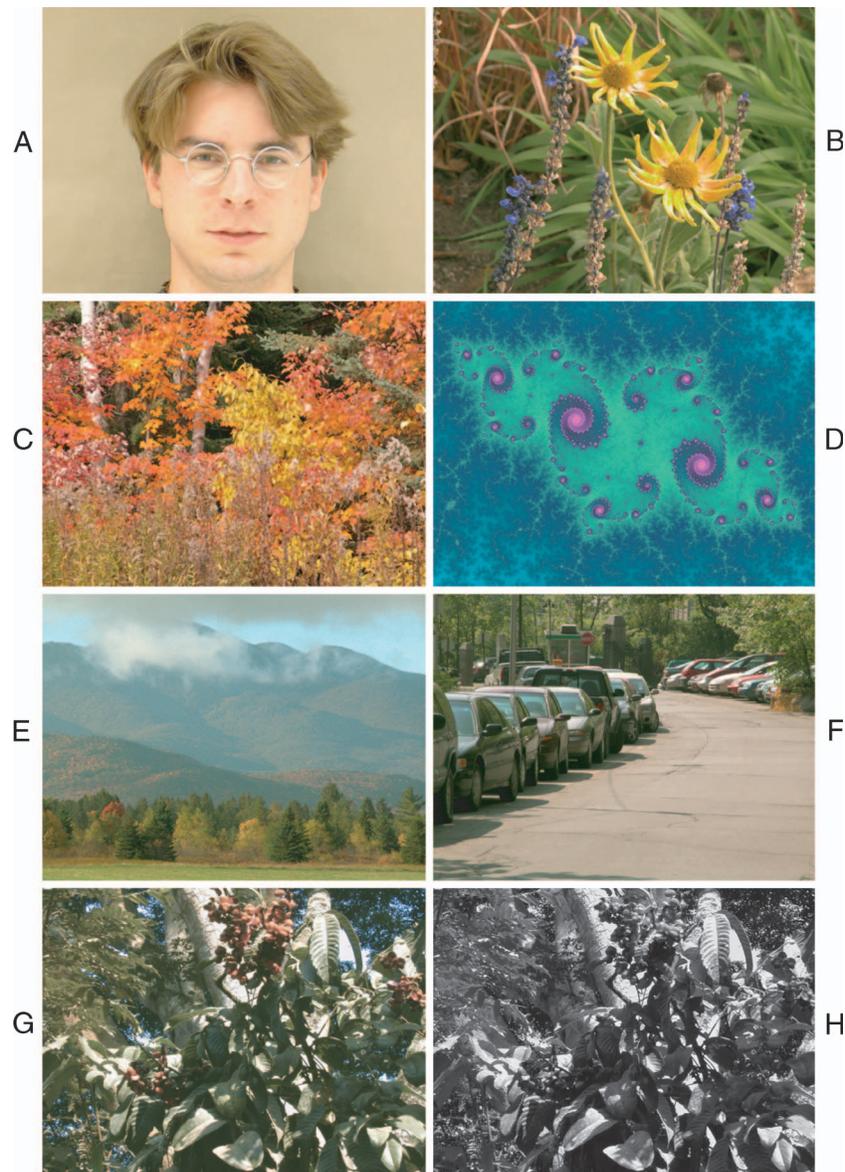


Figure 1. Example images. Colored Face (A), Flower & Animal (B), Forest (C), Fractals (D), Landscape (E), Man-Made (F), and Rainforest (G). The grayscale version of the image in panel G is shown in panel H.

subject's pupil position. We measured eye positions at a sampling rate of 250 Hz. Saccades and fixations were defined based on four parameters: a saccade was detected if the acceleration exceeded  $8000^\circ/\text{s}^2$ , the velocity was higher than  $30^\circ/\text{s}$ , a distance of at least  $0.1^\circ$  was covered, and a minimum duration of 4 ms exceeded.

Before each block of stimuli, the eye-tracking system was calibrated using a nine-point calibration: nine fixation points appeared successively on the screen in random order, and subjects were asked to fixate them. This procedure was continued until a mean calibration error below 0.4 degrees of visual angle was reached, and the eye with lower error was then selected for monocular recording.

Presentation computer and monitor, eye-tracker and recording computer were positioned in the same darkened

room. The experimenter was present in the room for the duration of the experiment.

## Experimental design

Subjects' eye positions were continuously recorded while they freely explored the presented images. In order to minimize any instruction-related bias, we instructed the subjects to "study the images carefully." Each image was presented for 6 seconds. Between two consecutive stimuli, a fixation point was displayed at the center of the screen. The experimenter manually prompted presentation of the next stimulus after the subject had fixated this point. In the following, we will use the term "trial" to refer to the

fixations made by one subject on one image of a given category and condition.

The experiment was conducted in two sessions of 4 blocks each (3 blocks with 50 trials and then a last block with 41). The order of presentation was randomized for each subject, with the constraint that no image was presented in both conditions (grayscale and color) within the same session. The time between sessions was limited to a minimum of 11 days (with an average of 24 days) to minimize memory effects.

## Definition of features

We analyzed the influence of two luminance features and three color features on subjects' fixation behavior: luminance contrast, texture contrast (2nd order luminance contrast), saturation, RG color contrast, and BY color contrast (RG and BY contrast, respectively). These features were chosen due to their neurophysiological relevance.

The *luminance contrast* of a fixation point is defined as the standard deviation of luminance in a region around the fixation, normalized by the mean luminance of the whole image (Reinagel & Zador, 1999). *Texture contrast* is the canonical extension of luminance contrast and is the standard deviation of luminance contrast of a patch divided by the mean luminance contrast of the whole image. Normalization by the patch mean yields only very small feature value differences but does not change the overall results. Therefore, we will report only the results using the normalization by the mean luminance or mean luminance contrast of the whole image.

The two *color contrasts* were defined solely as the standard deviation of the chromatic content of an image patch along the cardinal color axes. Unlike in the case of luminance-related features, we did not normalize by the mean color value of the image. Since DKL values range from  $-0.5$  to  $0.5$ , a symmetrical distribution of color values along any color axis of DKL space would lead to a mean value of 0. The mean therefore is not a good normalization factor.

In DKL color space, the saturation of a pixel is given by the absolute value of the pixel's chromatic content. The *saturation* in an image patch was defined as the mean saturation of all pixels in that patch.

Feature values are computed in an 81 pixel (approximately  $2.3^\circ$ ) square patch around a given pixel. We chose this size of image patch in line with earlier studies (Einhäuser & König, 2003; Frey et al., 2007). Alternative patch sizes, ranging from 41 to 161 pixels, were also used for all features but did not lead to any qualitative difference in the results.

## Feature analysis

In order to assess the influence of stimulus features on overt attention, we applied the following procedure, which

avoids the potential confound of “central bias” (see Frey et al., 2007; Tatler et al., 2005).

For each subject and stimulus, we define the *actual value* as the median of the feature values over all fixation locations on the stimulus. Each actual value was compared with a corresponding baseline that took into account potential biases in the subjects' eye positions (Figure 2, panel B). Control fixations were defined as all fixations of the same subject on all other images in the same category (e.g., Face or Landscape) and condition (colored or grayscale). Calculating the median of the feature values at the control locations on the actual image yields the *control value*. The actual value should differ from the control value if and only if the feature has an effect on overt attention. As these values were not normally distributed (Figure 2, panel C), we tested the significance of this difference using a non-parametric statistical test—the two-sided Kolmogorov–Smirnov test (KS-test). Significance values were Bonferroni corrected because of the multiple comparisons performed for each feature. Features for which actual and control distributions differed with  $p < .01$  were termed salient.

In order to compare the differences between actual and control values among different categories and conditions, we employ the receiver operating characteristic (ROC). This measure can be used to describe how well we can discriminate between fixated and non-fixated regions based on feature or saliency values. The theoretical ROC curve is the plot of the sensitivity (true positive rate) versus 1-specificity (false positive rate) for all possible threshold values. The area under the curve (AUC) can be interpreted as the probability to observe for example higher luminance contrast at fixated regions than at non-fixated regions, when we randomly select a pair of fixations (Faraggi & Reiser, 2002). Perfect discrimination will yield a value of 1.0, whereas chance level is at 0.5.

## Congruency of fixation locations between conditions and observers

The design of our study allows us to determine a general influence of color on overt attention by looking at the distribution of fixations. If fixations differ between the colored and grayscale version of the same image, then color information influences overt attention. If the fixation behavior of different observers becomes more similar in colored images, then we can assume a causal influence. To assess the congruency of fixations between conditions and observers, we use information theoretic measure, the Kullback–Leibler divergence (KL-divergence; Kullback & Leibler, 1951), calculated according to Dayan and Abbott (2001) by

$$d_{\text{KL}} = \sum_{x,y} P(x,y) \log_2 \left( \frac{P(x,y)}{Q(x,y)} \right), \quad (1)$$

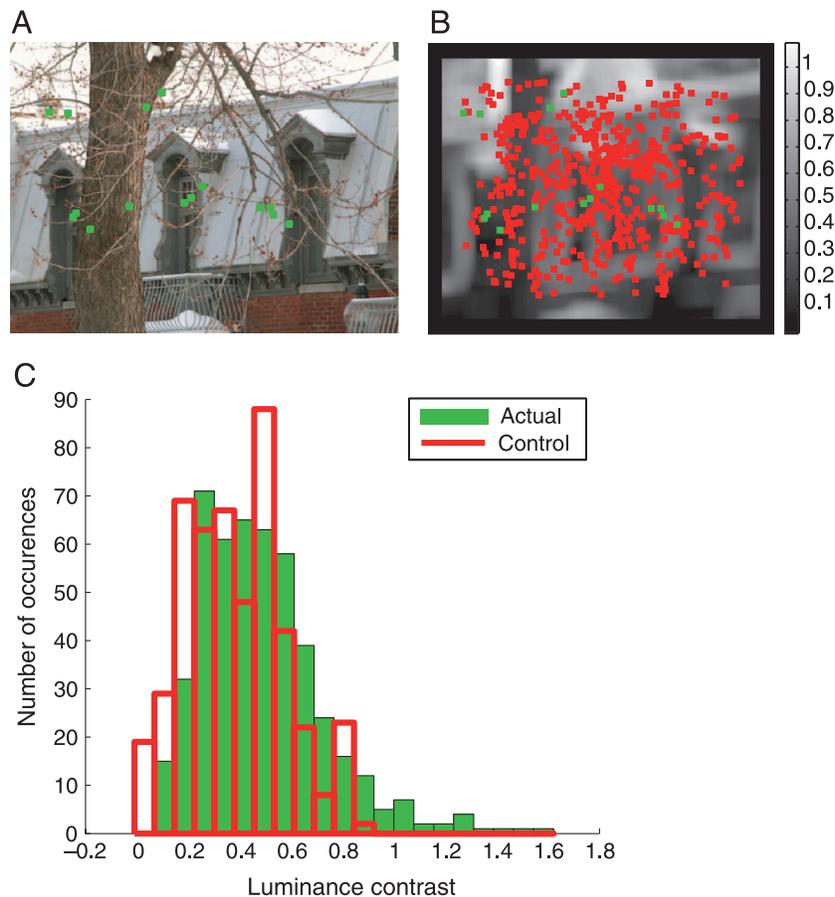


Figure 2. Feature analysis. (A) Measured fixation locations (green) of one subject on an image from the Man-Made category. The actual value is defined as the median feature value over all fixations of one subject on one stimulus. This image will be used in all further descriptions of statistical analyses. (B) Fixation locations (green) and corresponding control locations (red, see text for details) plotted on the luminance contrast map of this image. Control values are defined in an analogous manner. In our example the control value is somewhat higher than the actual value (0.37 and 0.35, respectively). (C) The distribution of actual (green bars) and control (opaque bars with red edge) luminance contrast for all subjects and images of Man-Made objects. The KS-test indicates that these two distributions are significantly different with  $p < .01$ . The ROC AUC value is 0.62. For presentation, the distributions are binned using 20 bins.

using point-wise multiplication and division. It can be regarded as a distance between two probability distributions  $\mathbf{P}$  and  $\mathbf{Q}$ , although it is not a real distance measure since it is not symmetric. Higher KL-divergence values indicate a bigger difference between fixation maps.

To determine the inter-observer congruency, we define two types of fixation probability distributions for each subject and image. At each fixation location we convolve a unit impulse with a 2D Gaussian with half-width at half-height of  $1^\circ$  visual angle. The size of the Gaussian is chosen in accordance with previous studies (Le Meur et al., 2007; Peters et al., 2005) and takes into account the precision of the eye-tracker. We divide this map by the sum of its entries to obtain the probability distribution. The first probability map is obtained from the fixations of a given subject and the second map is created using the fixations of all other subjects (Figure 3).

To determine the congruency between fixation locations on colored and grayscale versions of the same image seen by the same subject, we employ an identical approach.

The first probability map is created using fixations on the image presented in the colored condition, the second map using the fixations from the grayscale condition. For calculation of KL-divergence, we always used a maximum of 18 fixations per image. We chose this value because we obtained at least 18 fixations in about 3/4 of all trials.

## Saliency map

The saliency map model of bottom-up visual attention (Koch & Ullman, 1985) has been implemented in several different ways. One of the most prominent implementations was developed by Itti et al. (1998), and the source code for a software package including this model is freely available under GNU public license. We used this package, the iNVT C++ saliency toolkit (<http://ilab.usc.edu/toolkit/home.shtml>; build: 3.1 June 2007), with all parameters set to default values.

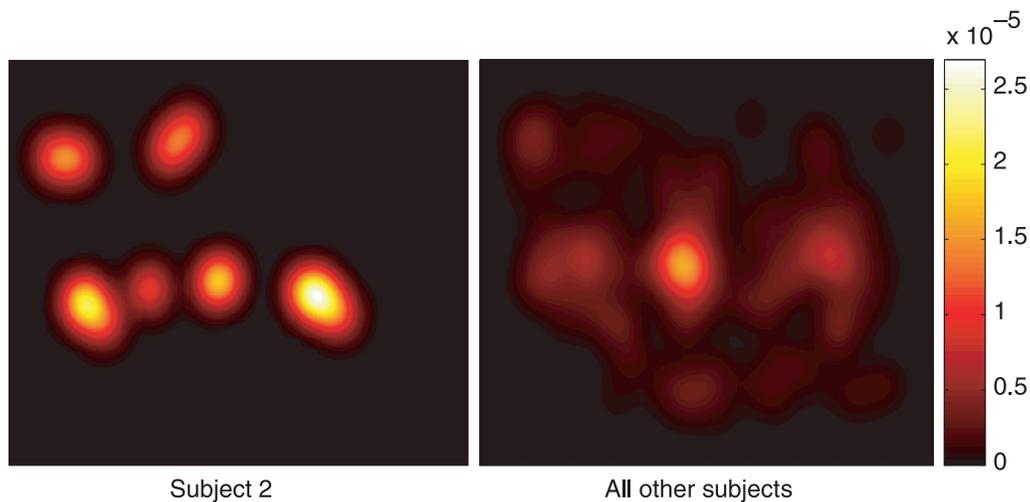


Figure 3. Calculation of congruency between observers. We create a fixation probability map for each subject (left) as well as for all other subjects (right). These two probability distributions are then compared using the Kullback–Leibler divergence. In this example, the KL-divergence is 20.09 bits.

The saliency toolkit allows calculation of a saliency map for each stimulus. Each element of such a map is a scalar value that indicates how salient, i.e., interesting to look at, it is. In order to determine how well the saliency measure can discriminate between fixated and non-fixated regions, we used ROC analysis. This procedure is equivalent to the ROC analysis of stimulus features outlined above.

## Control experiment

We presented 5 additional subjects with the same images. However, these subjects saw the images twice in the same condition (colored or grayscale). The time between sessions was 14 days. In the control experiment, images were presented on a 21-in. CRT monitor (NEC MultiSync FE2111, NEC; CIE coordinates of the phosphors: red 0.626/0.341, green 0.273/0.587, blue 0.151/0.065) at 100-Hz vertical refresh rate. This monitor was calibrated in order to achieve the same gamma, white point, and maximum luminance as in the main experiment. Eye movements were recorded using the EyeLink CL system (SR Research, Ontario, Canada) using the same parameters as in the main experiment.

## Results

### Eye movements of human subjects

The main question of this article is how color influences overt attention and what aspects of color are salient. To assess the influence on eye movements, we analyze two

different properties of human fixation behavior. First, we analyze whether color information changes the fixation behavior of individual subjects when compared to grayscale presentation. Second, we analyze whether color information attracts observers gaze. If color information really influences overt attention, then subjects' fixations should be directed toward more similar locations in colored images compared to grayscale images. For both analyses, we use the KL-divergence, an information theoretic measure for the distance between two probability distributions.

A necessary prerequisite for showing an influence of color on overt attention is that fixation locations change between colored and grayscale versions of the same image. Therefore, we determined each subjects' congruency of fixation locations between colored and grayscale presentation. These values are compared to the congruency of fixation locations between the same conditions of each image, as determined by our control subjects.

If a subject is presented with the same grayscale *Fractal* or *Man-Made* image in both sessions, the congruency of fixation locations between these two sessions is higher (lower KL-divergence) than if he or she is presented with the same image in different conditions. The same holds for the repeated presentation of colored *Landscape*, *Man-Made*, or *Rainforest* images (Figure 4, panel A). For these four categories, we can assume an influence of color on overt attention. In *Face*, *Flower*, and *Forest*, we find no significant differences between our different congruency measures. In *Face*, we find the highest congruency between fixation locations, with values as low as 6.5 bits (an example for fixation distributions leading to this value can be found in Figure 4, panel B). This is expected, since we know that faces are scanned in a stereotypical manner. In all other categories, we find significantly lower congruency values.

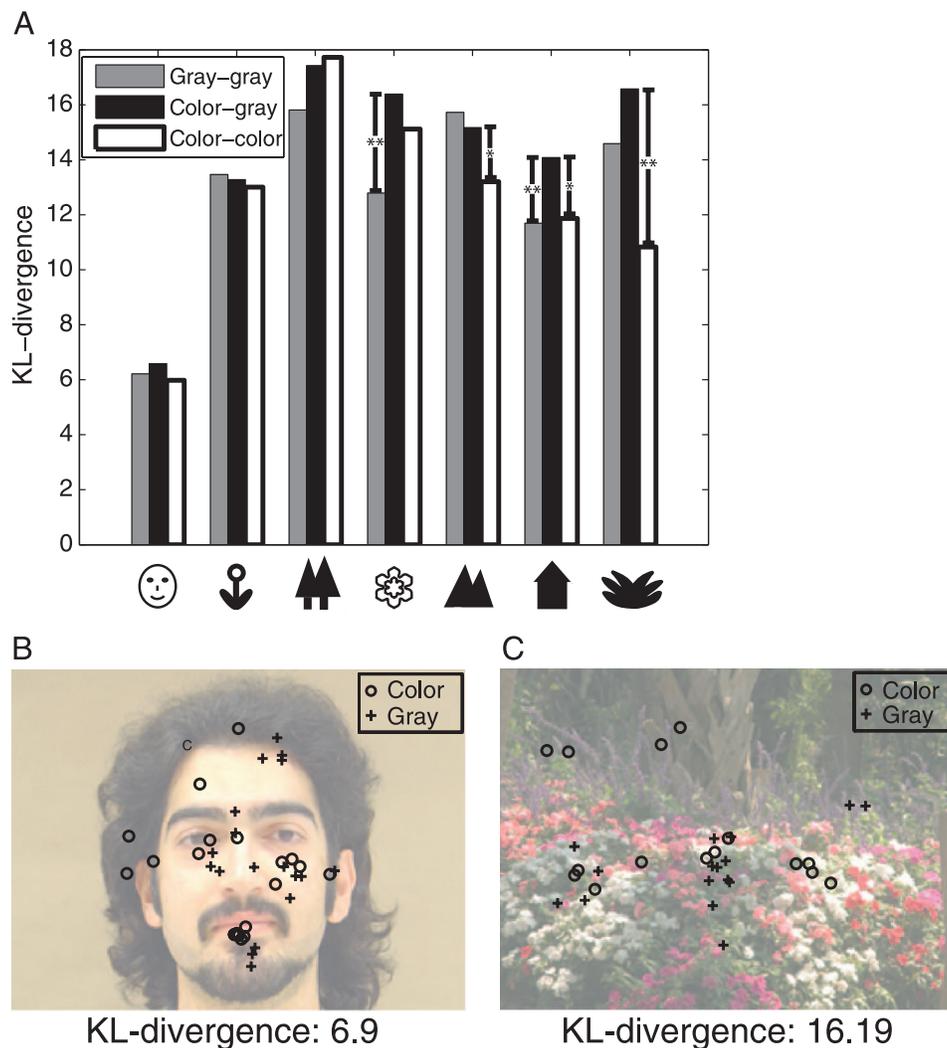


Figure 4. Similarity of fixation locations in colored and grayscale conditions. (A) Mean KL-divergence (with *SEM*) between fixation locations on colored and grayscale images. Each subject of the main experiment (black bars) saw each image in both conditions. For comparison, we plotted the KL-divergence for subjects who saw the same image twice in colored (white bars) or grayscale (gray bars). The icons on the x-axis represent the categories: Face, Flower, Forest, Fractal, Landscape, Man-Made, and Rainforest. (B and C) Example fixation distributions on colored (circle) and grayscale (cross) images, yielding low (B) and high (C) KL-divergence values. A high KL-divergence value corresponds to a low congruency between fixation locations.

Our results indicate that there is an influence of color on overt attention in *Fractal*, *Landscape*, *Man-Made*, and *Rainforest*. The nature of this influence will be analyzed in the following sections.

As a next step, we determine the congruency between observers. We first assess the congruency between observers in colored images and then compare this to the congruency in grayscale images. If color information causally influences overt attention, then we expect inter-observer congruency to be (significantly) higher in colored images. For colored images, the inter-observer congruency differs strongly between the different categories. As expected from earlier studies, we find a high inter-observer congruency for *Face* (KL-divergence 7.53; Figure 5, panel A). In all other categories the congruency

is significantly lower, with the highest KL-divergence in *Forest* (17.9; Figure 5, panel A). In certain categories with low inter-observer congruency, like *Forest* or *Landscape*, the saliency of image locations or objects seems to depend more on subjective appraisal of the individual subject. This also indicates that there are certain categories in which it is more difficult to predict where a subject is going to fixate.

Comparing colored and grayscale presentations, we find no significant differences in inter-observer congruency except for *Rainforest*. In *Rainforest* there is a significantly decreased KL-divergence in colored images compared to grayscale images ( $p < .01$ , KS-test with Bonferroni correction; Figure 5, panel B), meaning that the color information in *Rainforest* images leads subjects to fixate

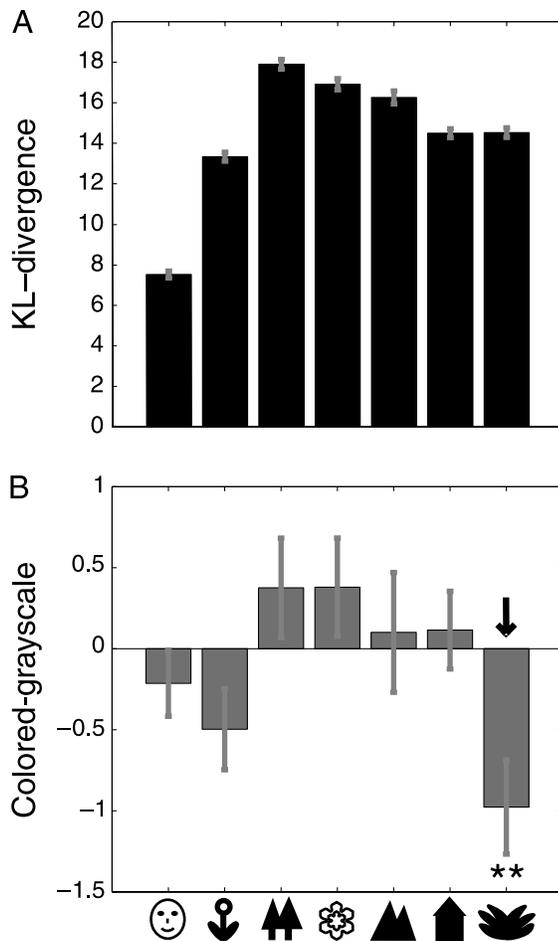


Figure 5. Inter-observer congruency. (A) Mean KL-divergence (with SEM) between fixation locations of different observers on colored images. Low KL-divergence values indicate a high congruency of fixation locations. (B) Difference in KL-divergence between colored and grayscale images. Values smaller than 0 indicate a higher congruency between observers in colored images.

more similar locations. There are no significant differences in all other categories. In *Flower*, we find a tendency for higher congruency between observers in colored images. In *Face*, *Landscape*, and *Man-Made*, the congruency remains virtually unchanged. We find a tendency for lower congruency between observers in colored images in *Forest* and *Fractal* (Figure 5, panel B).

Although the difference is not significant, it is surprising that in two of the categories color information makes subjects look at less similar locations. There are two possible explanations for this effect. Either color information is not attracting fixations, or it increases variability in fixation by making additional locations in the image salient. This question will be dealt with in the next section about image features, where we will also analyze which aspect of color increases inter-observer congruency in *Rainforest*.

## Stimulus features

### Luminance features

Fixated regions differ from non-fixated ones with respect to several image features. In grayscale images, luminance contrast and texture contrast are features that allow, to some extent, to predict fixation locations. Since color information changes subjects' gaze, it is thus possible that available color information alters the saliency of image features.

We first analyze whether features that are salient in grayscale images are also salient when we add color information. This is done by comparing feature values at fixated locations to subject-specific control locations using the KS-test and ROC analysis, in both colored and grayscale images.

In grayscale images, luminance contrast is salient ( $p < .01$ , KS-test with Bonferroni correction) in all categories except *Forest* and *Fractal*. Texture contrast is only salient in images of *Man-Made* and *Landscape* (Figure 6, gray bars). In colored images, we find exactly the same pattern. However, the ROC AUC values for these features are somewhat lower than in grayscale images (Figure 6, black bars). In order to quantify the similarity between attended luminance features in colored and grayscale images, we calculated the correlation between these features at fixated locations in all trials. The correlation between luminance contrast in colored and grayscale images across all trials is 0.9, while it is 0.71 for texture contrast (Figure 6B and 6C, respectively).

In conclusion, we find a very high similarity between colored and grayscale images when looking at the feature values and ROC AUC values for luminance and texture contrast. If a luminance feature is salient in grayscale images, it is also salient in colored images. This means that available color information does not significantly change the saliency of luminance features.

### Color features

Relatively little is known about which color features attract overt attention in different categories of images, as virtually all published studies with different categories of stimuli used coarse color features. Here we examined the saliency of three neurophysiologically plausible features in colored images: saturation, RG, and BY contrast.

RG and BY contrast are salient in *Face*, *Landscape*, *Man-Made*, and *Rainforest* categories (Figure 7, panel A). The strongest influence of both features is in *Rainforest*. The ROC AUC values for these two features are very similar for all categories. Saturation is a salient feature in *Flower*, *Forest*, and *Rainforest* (Figure 7, panel A). Again, the strongest influence is found in *Rainforest* images.

Color contrasts are salient especially in those categories in which at least one luminance feature is also salient. Saturation and color contrasts differ in saliency in the categories *Face*, *Flower*, *Forest*, *Landscape*, and *Man-Made*.

This shows that saturation and color contrasts selectively influence overt attention in different categories.

It is possible that color features are correlated with other image features, and we next made an analysis to reveal such correlations. We define a color feature as *originally*

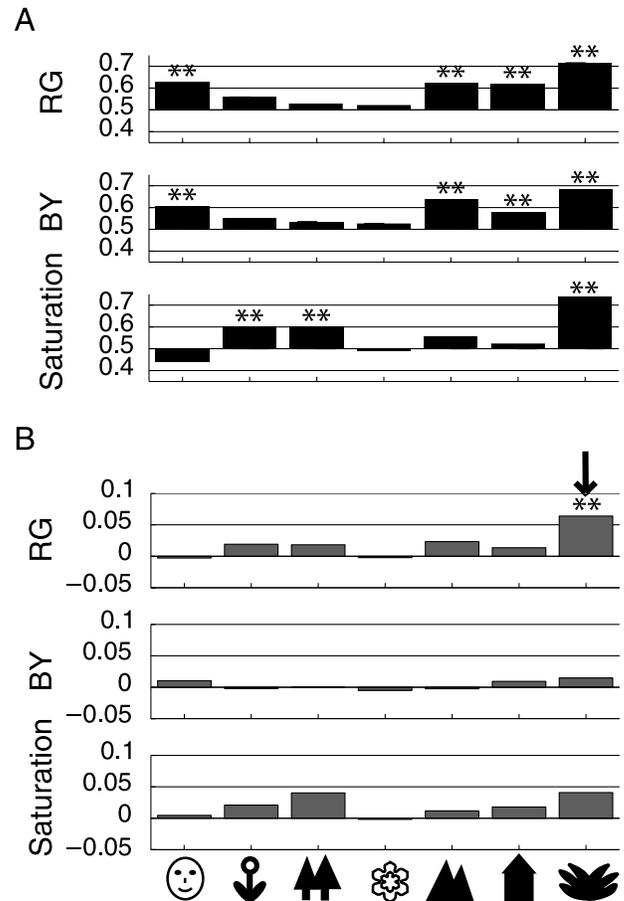
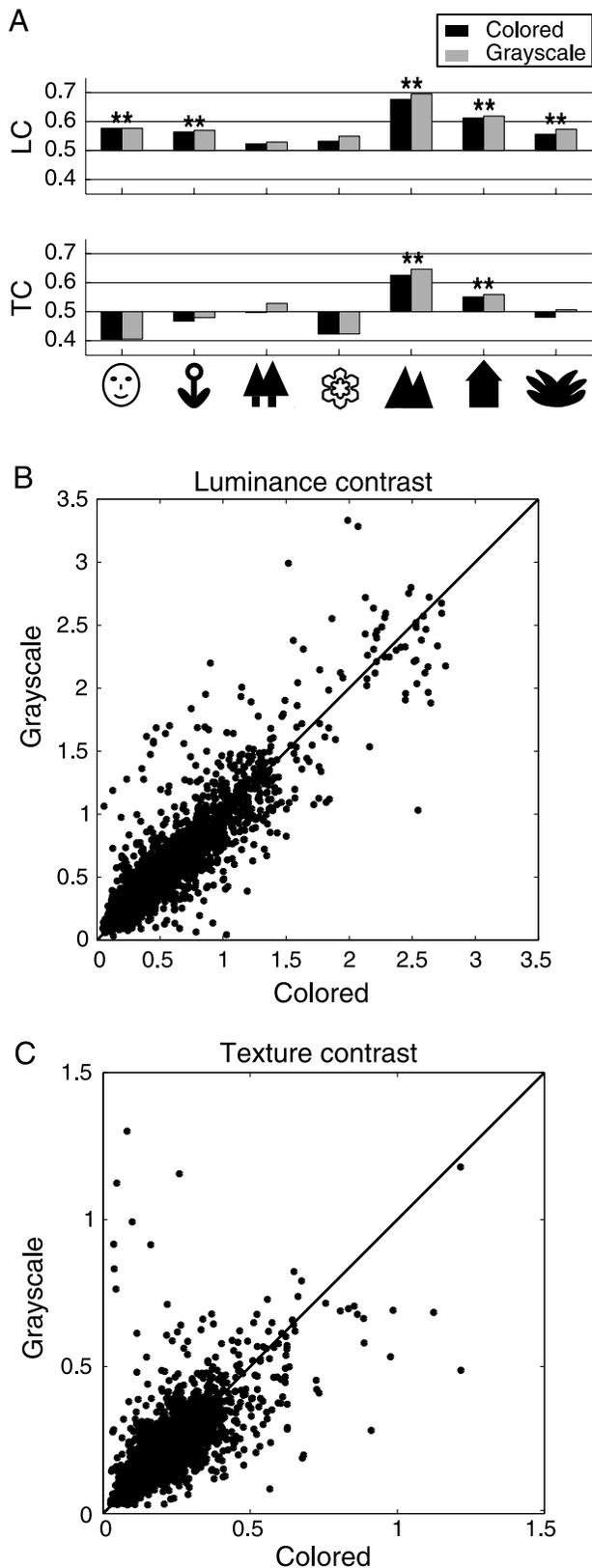


Figure 7. Color features. (A) ROC AUC for features RG contrast, BY contrast, saturation in colored images. Two asterisks indicate a significant difference between fixated and control locations ( $p < .01$ , KS-test with Bonferroni correction). (B) Difference in ROC AUC values for the same color features between fixations made in the colored condition and fixations made in the grayscale condition. Values greater than 0 indicate that a better discrimination between fixated and non-fixated image locations can be made using the color feature calculated at fixations measured in the colored condition.

Figure 6. Luminance features. (A) ROC AUC for features luminance contrast (upper panel) and texture contrast (lower panel). The black bars represent luminance features in colored images, the gray bars grayscale images. Two asterisks indicate a significant difference between feature values at actual and control locations ( $p < .01$ , KS-test with Bonferroni correction). (B) Luminance contrast feature values at fixated locations for all colored and grayscale images, all categories pooled. The correlation coefficient is  $r = .9$ . Least squares linear regression analysis returns a slope of 0.99. (C) Texture contrast values at fixated locations for all colored and grayscale images. The correlation coefficient is  $r = .71$ . Least squares linear regression analysis returns a slope of 0.95.

*salient* if it guides our attention not only by virtue of such a coincidental correlation. To assess the original saliency of each color feature, we created a distribution of comparison fixation locations for each image, made up of fixations made on that same image in the grayscale condition. Color feature values (of the colored stimuli) were then calculated at these comparison locations. Next we calculated the ROC AUC value for each color feature and stimulus category. Finally, we subtracted these ROC AUC value for comparison fixations from those of the actual colored condition fixations. This allows a comparison among different features, which is not possible with feature value differences due to the differing ranges of values of each feature.

Only the *Rainforest* category yields a significantly higher RG contrast at fixation locations measured in the colored condition compared to fixation locations from the grayscale condition ( $p < .01$ , KS-test with Bonferroni correction). The difference between the ROC AUC values for colored and grayscale presentation is 0.065 (Figure 7, panel B). There are differences in other categories, and although these are not significant, we chose a difference of 0.02 in ROC AUC values as a lower limit for assuming at least some effect of colored presentation.

Summarizing the analyses of color features (Figure 9), we find that RG and BY contrasts are salient in colored *Face*, *Landscape*, *Man-Made*, and *Rainforest* images. In *Face*, however, we find no difference between RG contrast ROC AUC values arising from colored and grayscale image presentations. Therefore, it is very likely that RG contrast is not truly salient in this category, but rather correlated with other luminance-defined features. In the remaining three categories (*Landscape*, *Man-Made*, and *Rainforest*), RG contrast is originally salient. BY contrast values do not differ between grayscale and colored presentation in *Landscape* and thus does not seem to be originally salient in this category.

Saturation is salient in colored *Flower*, *Forest*, and *Rainforest* images. Subtracting possible correlations with other features leaves saturation originally salient in all three categories. Overall, the color features analyzed here influence overt attention selectively in different image categories. Furthermore, analyzing only one chromaticity feature is probably not sufficient to reveal the influence of color on overt attention.

With regard to subjects' fixation behavior, it seems that the elevated congruency between observers in colored *Rainforest* images may be a consequence of the saliency of RG contrast. No other feature is salient after subtracting

any possible correlation with other features, suggesting that RG contrast is truly attracting attention in *Rainforest* stimuli. In *Forest* no color contrast and in *Fractal* no color feature analyzed is salient. These two are the categories in which we find a decreased inter-observer congruency in colored compared to grayscale images. This speaks in favor of the first possible explanation for this effect, namely that color features are not salient in these categories.

## Saliency map model

The goal of this study is to analyze the influence of (bottom-up) color features on overt attention. Since the saliency map is one of the leading models of bottom-up visual attention, we examined its performance in predicting fixation locations of human subjects. As above, we begin by analyzing whether predicted fixation locations differ between colored and grayscale versions of the same image. Next we determine how well the saliency map model predicts human fixations. Finally, to analyze the influence of color information, we compare how precise the predictions of the saliency map model are in colored and grayscale images.

In human subjects the fixation locations in the colored condition of an image differ from those in the grayscale condition. It is only in *Face* that this effect is not very pronounced. To assess the influence of color on the saliency map, we applied the same analysis as used in the between-condition comparison of human fixation maps. The mean KL-divergence values between saliency maps for colored and grayscale versions of the same image are very small, ranging from 0.02 in *Forest* to 0.05 in *Landscape* (Table 1). Thus, the saliency map model is not very strongly influenced by the presence of color information.

Previous studies have shown that saliency map models can predict human fixations well above chance. We compared the prediction performance of the standard model for our 7 categories of images by using the area underneath the ROC curve to quantify how well fixated and non-fixated regions can be discriminated based on their saliencies. The ROC AUC in colored images is highest for *Landscape* (Figure 8, panel A). In *Flower*, *Man-Made*, and *Rainforest*, we also find values higher than 0.65. These high AUC values indicate that it is possible to discriminate well between fixated and non-fixated regions

Face	Flower	Forest	Fractal	Landscape	Man-Made	Rainforest
0.04	0.04	0.02	0.03	0.05	0.03	0.03

Table 1. Influence of color on the saliency map model: mean KL-divergence between saliency maps for colored and grayscale versions of the same image.

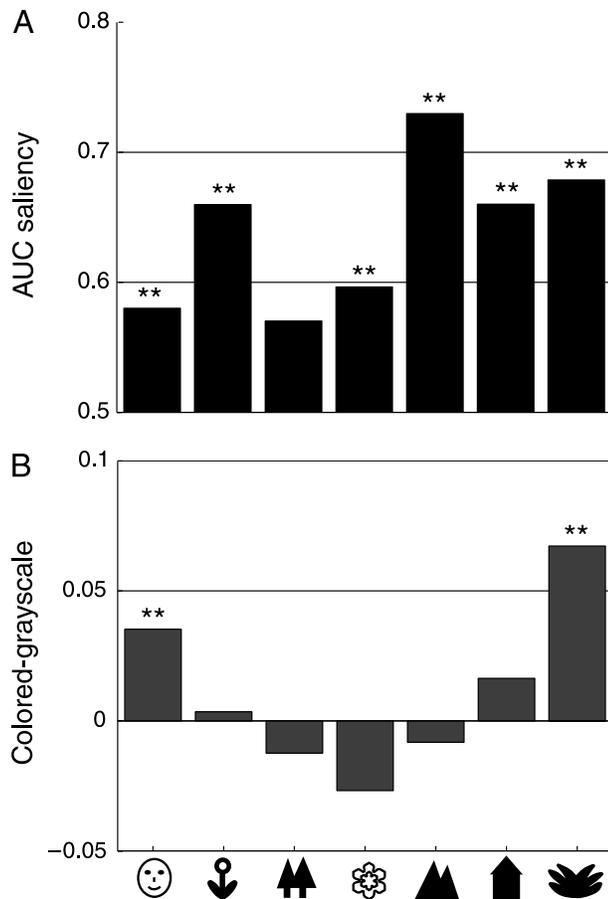


Figure 8. Saliency map model. (A) ROC AUC for discrimination between fixated and non-fixated image locations based on saliency values. Two asterisks indicate that saliency at fixated locations differs significantly from that at control locations ( $p < .01$ , KS-test with Bonferroni correction). (B) Difference in ROC AUC for saliency between colored and grayscale images. Values higher than 0 indicate an improvement in model performance in colored images.

based on saliency values. In these four categories subjects often fixate points which have a high saliency as determined by the saliency map model. In *Forest*, it is possible to discriminate between fixated and non-fixated regions only slightly better than chance.

In conclusion, the AUC values we obtained are in the range of previously reported values. The model predicts fixations best in colored *Landscape* images. Intermediate performance is reached in *Flower*, *Man-Made*, and *Rainforest*, while its prediction is worst for *Face*, *Fractal*, and *Forest*.

In the case of human subjects, we found that color information attracts subjects' gaze to significantly more similar locations only for *Rainforest* images. It was only in this category that we found a feature that seems to causally attract overt attention. Is the saliency map model able to select these fixation locations in *Rainforest* images

as well? Are there other categories in which color improves model performance?

AUC values are significantly higher in colored compared to grayscale images in *Face* and *Rainforest* ( $p < .01$ , KS-test with Bonferroni correction; Figure 8, panel B). In these categories, color information improves the prediction performance of the saliency map. There is virtually no performance difference between colored and grayscale conditions for *Flower* images, while we find a slight improvement in *Man-Made*. We have three categories in which the model performance is reduced with color information available (*Forest*, *Fractal*, and *Landscape*), here color information deteriorates model performance.

The biggest reduction of performance is in the *Fractal* category, for which there is no statistical relation between image features. In *Face* and *Rainforest*, the AUC measure improves with available color information. These are also those categories for which an evolutionary advantage of trichromatic color vision is proposed. Trichromacy is advantageous for the perception of skin color signaling in Faces (Changizi, Zhang, & Shimojo, 2006). Although it may appear that the saliency map model makes use of the naturalness of color in these three categories, this effect seems to be mostly due to the changes in inter-observer congruency with available color information since the saliency maps do not change between colored and grayscale versions of the same image (Table 1).

## Summary

### *The influence of color on man and model*

Summarizing the above analyses, it becomes evident that there are two extreme categories of images. In *Rainforest*, color information improves all indices we analyzed and all color features are originally salient meaning that observers are strongly influenced by color in this category. In contrast, in *Fractal* no color feature is originally salient and inter-observer congruency as well as saliency map performance is worse in colored images. The other categories of images are somewhere in between these two extremes with *Forest* being very close to *Fractal*. With regard to the different color features, we can show that they differentially influence overt attention in the various categories.

## Discussion

In the present study, we demonstrate that there is a strong influence of color information on human overt attention. This manifests itself in the fact that fixation locations of human subjects differ between colored and grayscale versions of the same image. Interestingly, in two

categories (*Forest* and *Fractal*) the subjects' fixation locations become more dissimilar in colored images. It is only in *Rainforest* images that RG contrast makes subjects look at significantly more similar locations. When analyzing which aspects of color influence overt visual attention, we find that our chosen color features are selectively salient—the saliency of one color feature is not related to the saliency of other color features, and single color features are only salient in some categories (Figure 9).

## The influence of color in different categories

We find a strong general influence of color on overt attention in all categories except one. In *Face*, the fixation

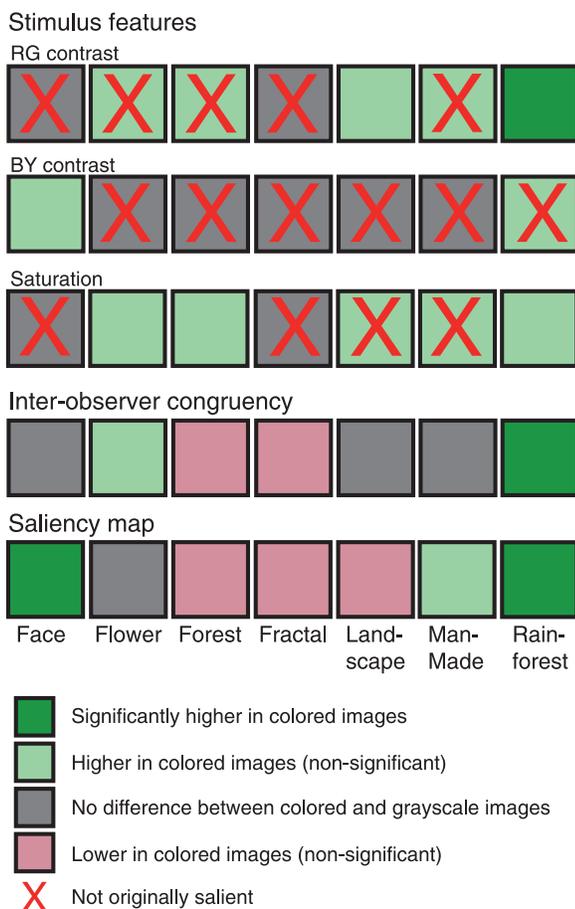


Figure 9. Influence of color on humans and saliency map. Dark green indicates significantly higher values in colored compared to grayscale images. Light green/red represents higher/lower values in colored images (non-significant). Gray indicates that there is no difference between colored and grayscale images with respect to a given measure. Red Xs label those categories in which a given color feature is not originally salient. This means that the feature is either not salient in colored images or its saliency is only due to a correlation with luminance defined features, as assessed by the AUC values for fixation on grayscale images.

locations in colored and grayscale version of *Face* stimuli are similar. Color information does not make subjects look at different locations. This parallels findings in face recognition tasks. Kemp, Pike, White, and Musselman (1996) have found that color is not a diagnostic feature for face recognition, unless shape information is degraded. These authors argue that it is rather shape cues (e.g., shape from shading) that are diagnostic for recognizing faces. Therefore, it is highly probable that grayscale face stimuli already contain sufficient information for face recognition. This is likely to be the same kind of information that draws subjects' visual attention to more or less the same locations in colored and grayscale *Face* images. Memory effects could affect our analysis of the general effects of color on overt attention. But we made sure to minimize any probable effect by balancing the condition of the first presentation of each image and by having a relatively long period between the two recording sessions.

Looking at the specific effects of color information on overt attention, we find two extreme categories of images. In *Fractal*, available color information makes subjects' fixation patterns more dissimilar. In addition, no color feature analyzed in this study is salient in the case of *Fractal* images. This indicates that the image features we analyzed do not influence subjects' overt attention in this category. *Fractal* is the only category in which image features are not related in a conventional way. For example, there is no correlation between luminance contrast and BY color contrast in this category, which is strongly present in all other categories that feature photographed images. It is likely that this lack of a natural statistical relation between features is one of the factors for the effects found in this category.

The other extreme category is *Rainforest*. It is the only category in which color information significantly improves the prediction of the saliency map model and the congruency between observers. All color features are salient in *Rainforest*, too, with an especially strong influence of RG contrast on overt attention. This influence remains even after a possible correlation with other features is removed. This finding is very interesting because it concurs with findings on primate trichromatic color vision (the ability to compare outputs of L and M cones; Nathans, 1999). It is in the *Rainforest* environment that trichromatic color vision evolved, and the cones of trichromatic primates are optimally tuned to detect food sources in such surroundings (Sumner & Mollon, 2000). Suppose then that the RG visual channel detects food sources like ripe fruits and edible young leaves that are rather sparsely distributed in the environment. In terms of the saliency map model, this yields a few high peaks in the RG feature map and these high peaks are then able to strongly contribute to the saliency map. Therefore, the finding that RG contrast is very salient in *Rainforest* is in agreement with the basic idea of neurobiologically plausible models.

In *Forest* and *Fractal* categories, the inter-observer congruency is reduced in colored images compared to grayscale images. The most likely explanation for this result is the lack of salience of color features in these categories—no color feature analyzed is salient in *Fractal* and no color contrast feature is salient in *Forest*. However, in the case of *Flower*, although the two color contrasts are not salient, there is still a tendency for higher inter-observer congruency in colored images. This rules out the very simple explanation that every time color contrasts are not salient, inter-observer congruency is diminished in colored images. Li and Lennie (2001) have shown that surface segmentation based on color variations is more successful than using corresponding brightness variations. This color variation property is immune to disruptions by chromatic noise, probably due to the capacity of the visual system to combine signals from a large region. The *Flower* category contains predominantly close-up images of colorful flowers, which means that they contain large homogeneously colored surfaces belonging to different parts of the depicted objects. In the case of such flower images, color information should be helpful in image segmentation. If subjects then fixate the centers of the segmented image region, this could be the reason that color contrasts are not salient in this category but rather that saturation is.

## Saliency map

The saliency map model exhibits good prediction performance in more than half of the categories. The ROC AUC values for colored images are at the upper end of the range of values reported in previous studies. In *Flower*, *Landscape*, *Man-Made*, and *Rainforest*, it is possible to discriminate well between fixated and non-fixated locations based on saliency values. In *Face*, *Forest*, and *Fractal*, the model predicts human fixation locations only slightly better than chance. Interestingly, in *Face*, we find the highest congruency between human observers. This indicates that there are certain features that attract the attention of the vast majority of subjects. Subjects predominantly fixate on eyes, nose, ears, and mouth—the saliency map model however is not capable of detecting these features. A neurobiologically plausible way to improve model performance in this category could be to incorporate knowledge about object or face processing that takes place in higher visual areas. Such an approach was taken by Cerf, Harel, Einhäuser, and Koch (in press). They showed that if faces are present in an image they are typically fixated within the first two fixations. Adding a simple face detection module improved the performance of the saliency map model dramatically in images containing faces. However, Cerf and colleagues did not use close-up images of faces as we used in our study. Hence, it is a difference to detect a face in a cluttered scene and scanning different parts of a

close-up shown face. So it is unclear whether this approach could enhance the performance of the saliency map model in our setting.

The other two categories in which the model performs poorly are *Flower* and *Fractal*. In these categories, the congruency between observers is lowest. It could well be that the more scattered fixation locations influence our performance measure in these two categories.

Concerning the influence of color on the saliency map model, we find that the saliency map for a colored stimulus does not differ from the map of its grayscale counterpart. We find this throughout all categories and images. This means that color does not influence the generation of the saliency map in the model we examined.

Nonetheless, the saliency map model predicts human fixations better in colored *Face* and *Rainforest*. The high congruency found between image saliency maps from color and grayscale images tells us that in grayscale images the saliency map already has high activity at locations fixated by human subjects in colored images. The improvement seen in saliency-based discrimination when color information is present means that subjects look at those locations more often in colored images than grayscale images, i.e., these regions are not salient for human subjects in grayscale images. One possible explanation for this effect could be a correlation of luminance and color features. If both features exhibit high feature values, the linear combination of feature maps means saliency increases at such locations. And indeed we find high correlation coefficients between luminance contrast and color contrasts in *Face* stimuli, with values around 0.6. We find almost exactly the same correlation coefficients at fixated and at control locations. However, the correlation coefficients in *Rainforest* are significantly lower. Therefore, it seems that a combined feature effect can explain the results in *Face* stimuli, but not in *Rainforest*. What causes this effect in *Rainforest* stimuli remains unclear.

## Is *Rainforest* special?

Based on the fact that trichromacy in primates evolved in rainforests, we expected that *Rainforest* would differ from other categories with respect to the saliency of color features. The RG color channel in particular should exhibit high saliency. Indeed, we found that all color features are highly salient in images of *Rainforest*. Moreover, RG contrast in *Rainforest* is the only feature for which we can exclude any possible correlation with other image features that are already present in grayscale images. Furthermore, it is only in *Rainforest* that the congruency between fixation locations of different observers is significantly higher in colored compared to grayscale images. This increased inter-observer congruency is most likely a direct consequence of RG contrast being originally salient.

All these data support our initial notion that *Rainforest* is a special category when examining the salience of color features. As expected, our results point toward a strong influence of RG contrast on overt attention in this category.

## Color space and experimental task

The cardinal color axes of the DKL color space are well suited to describe the preferred colors of neurons in LGN. This does not hold for cortical cells in V1 and V2 (Gegenfurtner & Kiper, 2003). For higher-level chromatic tasks like color appearance judgments, the cardinal color axes are irrelevant. However, we chose the DKL color space for several reasons. First, the color processing of the saliency-map model mimics the processing of the color-opponent cells in LGN. The DKL color space describes neuronal responses in LGN. Therefore, we could determine features similar to the saliency-map model. Second, we wanted to define color features independent of assumptions about the spatial scale of the stimulus. Color spaces suitable for higher-level tasks like CIE LUV are based on the 1931 CIE XYZ tristimulus values. This means that they are normally defined for a 2° field of view. However, there are homogeneously colored objects of up to 10° in our stimuli. Therefore, we did not analyze our data using other color spaces.

The subjects were given the task to “study the images carefully.” This imposes the danger that each subject pursues a different strategy. There are three reasons why we gave this type of task. First, in the critical comparison of previous psychophysical studies (e.g., Frey et al., 2007; Tatler et al., 2005), we have to use the same task as the studies before. Third, in a separate study (Betz, Kietzmann, Wilming, & König, in preparation), we demonstrate that this task instruction leads to a similar distribution of fixation points as more semantically involved tasks. However, this does not eliminate the need for future studies to compare different tasks.

## Color features

Studies in monkeys and humans have revealed that color can be used as an efficient bias during visual search (e.g., Bichot, Rossi, & Desimone, 2005; for a review, see Desimone & Duncan, 1995). The experimental task employed in this study is neutral with respect to stimulus features. In addition, the random presentation of different stimulus categories and conditions makes priming effects unlikely. Therefore, the lack of saliency of color features in several categories in our study does not preclude that they could be salient given a different task.

Up to now, there has been no systematic study of the influence of color features on overt attention in natural scenes. In published studies that employ different

categories of stimuli, color information was reduced to a single feature (e.g., Parkhurst et al., 2002) with no discrimination made between the two color processing pathways. Here, we describe color information using the neurophysiologically plausible features of saturation and color contrast in the RG and BY color channel. These features influence the firing of neurons in the retinogeniculate channels as well as parvocellular (and most probably also koniocellular) layers of LGN (Gegenfurtner & Kiper, 2003; Hendry & Reid, 2000).

Using these detailed color features in our analysis, we find different results than earlier studies. Contrary to Parkhurst et al. (2002), we find no influence of color features on overt attention in *Fractal* stimuli. In their analysis, the color feature had the highest “relative strength” of all features analyzed, i.e., should influence overt attention. This is clearly not the case in our study. Another difference concerns the relative influence of luminance and color features in their category “Buildings and City Scenes,” which is similar to our *Man-Made* category. Parkhurst and colleagues found a higher relative strength for luminance features than for color features. The same result was reported by Tatler et al. (2005) for images that could be attributed to the *Man-Made* category. Here we cannot replicate these findings. Instead, we find that RG contrast is more salient than luminance contrast.

There are three possible explanations for these differences. First, different studies employ different images. In virtually all studies (including ours), stimuli are selected on the basis of semantics. We simply cannot say whether these stimuli definitively capture the aspects that are crucial for a given category. A more rigorous approach would be to categorize natural stimuli “based on functionally relevant statistical properties” (Felsen & Dan, 2005), for example. Second, the feature extraction methods were completely different in these two studies. Parkhurst and colleagues combined the outputs of two color-opponent feature channels of the saliency map model into one color channel. In contrast, we used excitations along color-opponent channels of a neurophysiologically plausible color space to calculate our features. Third, our stimuli (except for *Face* stimuli) consisted of color-calibrated images, which are devoid of any color aberrations normally present in digital photographs.

Since the standardized categorization of stimuli is probably the most difficult aspect to deal with, we consider it outside the scope of this discussion. In this study, we have however dealt with the last two points by using different neurophysiologically plausible color features in analysis and artifact-free stimuli in our experiments. Analyzing the influence of these color features, we found two extreme categories of images, namely *Fractal* and *Rainforest*. While color impairs all analyzed indices in the first, it improves all in the latter. In the other categories, color features are selectively salient. This shows that the influence of color on overt attention

depends on the type of image. Also, it is crucial to analyze neurophysiologically relevant color features for quantifying the influence of color on attention.

## Acknowledgments

We thank Cliodhna Quigley for helpful comments on drafts of this manuscript and Alper Acik for providing the Face stimuli.

Commercial relationships: none.

Corresponding author: Hans-Peter Frey.

Email: hfrey@uos.de.

Address: Institute of Cognitive Science, Albrechtstrasse 28, 49076 Osnabrück.

## References

- Bichot, N. P., Rossi, A. F., & Desimone, R. (2005). Parallel and serial neural mechanisms for visual search in macaque area V4. *Science*, *308*, 529–534. [[PubMed](#)]
- Buswell, G. T. (1935). *How people look at pictures: A study of the psychology of perception in art*. Chicago: University of Chicago Press.
- Carmi, R., & Itti, L. (2006). Visual causes versus correlates of attentional selection in dynamic scenes. *Vision Research*, *46*, 4333–4345. [[PubMed](#)]
- Cerf, M., Harel, J., Einhäuser, W., & Koch, C. (in press). Predicting human gaze using low-level saliency combined with face detection. In A. McCallum (Ed.), *Advances in neural information processing systems (NIPS)*.
- Changizi, M. A., Zhang, Q., & Shimojo, S. (2006). Bare skin, blood and the evolution of primate colour vision. *Biology Letters*, *2*, 217–221. [[PubMed](#)] [[Article](#)]
- Dayan, P., & Abbott, L. F. (2001). *Theoretical neuroscience: Computational and mathematical modeling of neural systems*. Cambridge, MA: Massachusetts Institute of Technology Press.
- Derrington, A. M., Krauskopf, J., & Lennie, P. (1984). Chromatic mechanisms in lateral geniculate nucleus of macaque. *The Journal of Physiology*, *357*, 241–265. [[PubMed](#)] [[Article](#)]
- Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*, 193–222. [[PubMed](#)]
- Dominy, N. J., & Lucas, P. W. (2001). Ecological importance of trichromatic vision to primates. *Nature*, *410*, 363–366. [[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](#)]
- Faraggi, D., & Reiser, B. (2002). Estimation of the area under the ROC curve. *Statistics in Medicine*, *21*, 3093–3106. [[PubMed](#)]
- Felsen, G., & Dan, Y. (2005). A natural approach to studying vision. *Nature Neuroscience*, *8*, 1643–1646. [[PubMed](#)]
- Frey, H. P., König, P., & Einhäuser, W. (2007). The role of first- and second-order stimulus features for human overt attention. *Perception & Psychophysics*, *69*, 153–161. [[PubMed](#)]
- Gegenfurtner, K. R., & Kiper, D. C. (2003). Color vision. *Annual Review of Neuroscience*, *26*, 181–206. [[PubMed](#)]
- Gottlieb, J. P., Kusunoki, M., & Goldberg, M. E. (1998). The representation of visual salience in monkey parietal cortex. *Nature*, *391*, 481–484. [[PubMed](#)]
- Hendry, S. H., & Reid, R. C. (2000). The koniocellular pathway in primate vision. *Annual Review of Neuroscience*, *23*, 127–153. [[PubMed](#)]
- Hoffman, J. E., & Subramaniam, B. (1995). The role of visual attention in saccadic eye movements. *Perception & Psychophysics*, *57*, 787–795. [[PubMed](#)]
- Horwitz, G. D., & Newsome, W. T. (1999). Separate signals for target selection and movement specification in the superior colliculus. *Science*, *284*, 1158–1161. [[PubMed](#)]
- Ignashchenkova, A., Dicke, P. W., Haarmer, T., & Thier, P. (2004). Neuron-specific contribution of the superior colliculus to overt and covert shifts of attention. *Nature Neuroscience*, *7*, 56–64. [[PubMed](#)]
- Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews, Neuroscience*, *2*, 194–203. [[PubMed](#)]
- 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*, 1254–1259.
- Kemp, R., Pike, G., White, P., & Musselman, A. (1996). Perception and recognition of normal and negative faces: The role of shape from shading and pigmentation cues. *Perception*, *25*, 37–52. [[PubMed](#)]
- Kienzle, W., Wichmann, F. A., Schölkopf, B., & Franz, M. O. (2007). A nonparametric approach to bottom-up visual saliency. In B. Schölkopf, J. Platt, & T. Hoffman (Eds.), *Advances in neural information processing systems (NIPS)* (vol. 19, pp. 689–696). Cambridge, MA: MIT Press.

- Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: Towards the underlying neural circuitry. *Human Neurobiology, 4*, 219–227. [PubMed]
- Komatsu, H., Ideura, Y., Kaji, S., & Yamane, S. (1992). Color selectivity of neurons in the inferior temporal cortex of the awake macaque monkey. *Journal of Neuroscience, 12*, 408–424. [PubMed] [Article]
- Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *Annals of Mathematical Statistics, 22*, 79–86.
- Kustov, A. A., & Robinson, D. L. (1996). Shared neural control of attentional shifts and eye movements. *Nature, 384*, 74–77. [PubMed]
- Le Meur, O., Le Callet, P., & Barba, D. (2007). Predicting visual fixations on video based on low-level visual features. *Vision Research, 47*, 2483–2498. [PubMed]
- Li, A., & Lennie, P. (2001). Importance of color in the segmentation of variegated surfaces. *Journal of the Optical Society of America A, Optics, Image Science, and Vision, 18*, 1240–1251. [PubMed]
- Maioli, C., Benaglio, I., Siri, S., Sosta, K., & Cappa, S. (2001). The integration of parallel and serial processing mechanisms in visual search: Evidence from eye movement recording. *European Journal of Neuroscience, 13*, 364–372. [PubMed]
- Mannan, S. K., Ruddock, K. H., & Wooding, D. S. (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]
- Mazer, J. A., & Gallant, J. L. (2003). Goal-related activity in V4 during free viewing visual search. Evidence for a ventral stream visual salience map. *Neuron, 40*, 1241–1250. [PubMed] [Article]
- Nathans, J. (1999). The evolution and physiology of human color vision: Insights from molecular genetic studies of visual pigments. *Neuron, 24*, 299–312. [PubMed] [Article]
- Olmos, A., & Kingdom, F. A. A. (2004). *McGill calibrated colour image database*. Retrieved from <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. (2004). Texture contrast attracts overt visual attention in natural scenes. *European Journal of Neuroscience, 19*, 783–789. [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]
- Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annual Review of Neuroscience, 13*, 25–42. [PubMed]
- Reinagel, P., & Zador, A. M. (1999). Natural scene statistics at the centre of gaze. *Network, 10*, 341–350. [PubMed]
- Sumner, P., & Mollon, J. D. (2000). Catarrhine photopigments are optimized for detecting targets against a foliage background. *Journal of Experimental Biology, 203*, 1963–1986. [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]
- 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] [Article]
- Troscianko, T., Párraga, C. A., Leonards, U., Baddeley, R. J., Troscianko, J., & Tolhurst, D. J. (2003). Leaves, fruit, shadows, and lighting in Kibale Forest, Uganda. *Perception, 32* (EVP Abstract Supplement).
- von Helmholtz, H. (1867). *Handbuch der physiologischen Optik*. Hamburg: Verlag Leopold Voss.
- Zeki, S. (1983). The distribution of wavelength and orientation selective cells in different areas of monkey visual cortex. *Proceedings of the Royal Society of London B: Biological Sciences, 217*, 449–470. [PubMed]