

When viewing natural scenes, do abnormal colors impact on spatial or temporal parameters of eye movements?

Tien Ho-Phuoc

CEA, LETI/MINATEC, Grenoble, France, &
GIPSA Laboratory, CNRS, Grenoble, France



Nathalie Guyader

GIPSA Laboratory, CNRS, Grenoble, France



Frédéric Landragin

LATTICE Laboratory, CNRS and ENS, Paris, France



Anne Guérin-Dugué

GIPSA Laboratory, CNRS, Grenoble, France



Since Treisman's theory, it has been generally accepted that color is an elementary feature that guides eye movements when looking at natural scenes. Hence, most computational models of visual attention predict eye movements using color as an important visual feature. In this paper, using experimental data, we show that color does not affect where observers look when viewing natural scene images. Neither colors nor abnormal colors modify observers' fixation locations when compared to the same scenes in grayscale. In the same way, we did not find any significant difference between the scanpaths under grayscale, color, or abnormal color viewing conditions. However, we observed a decrease in fixation duration for color and abnormal color, and this was particularly true at the beginning of scene exploration. Finally, we found that abnormal color modifies saccade amplitude distribution.

Keywords: eye movements, natural scenes, color and abnormal color scenes, visual exploration, saliency modeling, visual attention

Citation: Ho-Phuoc, T., Guyader, N., Landragin, F., & Guérin-Dugué, A. (2012). When viewing natural scenes, do abnormal colors impact on spatial or temporal parameters of eye movements? *Journal of Vision*, 12(2):4, 1–13, <http://www.journalofvision.org/content/12/2/4>, doi:10.1167/12.2.4.

Introduction

When looking at natural scene images, our gaze is attracted to particular regions called *salient regions*. A lot of research has attempted to understand why some regions are salient regarding their statistical properties, using behavioral experiments with eye movement recording and/or computational models of the human visual system (Baddeley & Tatler, 2006; Buswell, 1935; Henderson & Hollingworth, 1999; Itti, Koch & Niebur, 1998; Le Meur, Le Callet, Barba, & Thoreau, 2006; Mannan, Ruddock, & Wooding, 1997; Marat et al., 2009; Parkhurst, Law, & Niebur, 2002; Privitera & Stark, 2000; Reinagel & Zador, 1999; Torralba, Oliva, Castelano, & Henderson, 2006; Yarbus, 1967). *Visual saliency* depends mainly on two factors: one is task independent and the other one is task dependent. The first one refers to bottom-up processes and is mainly driven by stimulus visual features (Koch & Ullman, 1985; Treisman & Gelade, 1980); the latter refers to top-down processes and is mainly driven by the task (Castelano, Mack, & Henderson, 2009; Henderson & Hollingworth, 1999; Yarbus, 1967). Most of the saliency models, also called visual attention models, simulate bottom-up processes to look for salient regions in visual stimuli; these regions are supposed to attract attention and,

hence, observers' gazes. Most computational models of visual attention are inspired by the *Feature Integration Theory* (FIT) of Treisman and Gelade (1980) and modeled on bottom-up processes. According to this theory, visual stimuli are first broken down into several feature maps such as intensity, color, and orientation; these features were shown to be encoded in the primary visual cortex and to evoke responses from different cortical cells (Hubel, Wiesel, & Stryker, 1977). A region is salient if its features differ from the surrounding features. The features are represented by separate feature maps, which are then combined to create a master saliency map. This map emphasizes salient regions. Besides intensity, color, and orientation visual features as mentioned in the FIT, there are several other salient visual features such as edges, spatial frequencies, and motion (Baddeley & Tatler, 2006; Wolfe & Horowitz, 2004). Usually, the following set of features—intensity, orientation, color, and spatial frequency—is taken into account in visual attention models to predict eye movements for exploring static scenes (Itti et al., 1998; Le Meur et al., 2006; Torralba et al., 2006). In this framework, it is accepted that color information contributes to fixation locations. As in the above models, several studies have shown the important role of color in visual attention (Frey, Honey, & König, 2008; Jost, Ouerhani, Wartburg, Müri, & Hügli, 2005;

Peters & Itti, 2008). It was revealed that in free viewing, gaze is attracted by color depending on the semantic category of the visual scene (Frey et al., 2008; Parkhurst et al., 2002). For example, Frey et al. (2008) used seven categories of stimuli: face, flower and animal, forest, fractal, landscape, man-made, and rainforest and found a difference in observers' fixation locations between color and grayscale scenes for the rainforest category. The role of color was also proved for different types of cognitive tasks. In recognition tasks, a saliency map taking into account color features better correlates with human fixations than a saliency map using only grayscale visual information (Jost et al., 2005). Similarly, a broad range of visual features such as orientation, intensity, color, flicker, motion, and their combinations is also tested for the prediction of eye movements in video games (Peters & Itti, 2008). The results of this study emphasize the role of color, whether it is used alone or combined with other visual features. Both in visual attention and also in object recognition, which might be associated with high-level vision, it was shown that color plays an important role making for faster object recognition, for example (see Tanaka, Weiskopf, & Williams, 2001 for a review).

However, a study by Tatler, Baddeley, and Gilchrist (2005) suggested that color has little involvement in fixation locations for natural images. In this study, the authors examined the ability of luminance, color, contrast, and edge to distinguish fixated locations from non-fixated locations. This was done according to different spatial frequencies and time. In both cases, the authors showed that color information correlates weakly with fixation locations compared to contrast or edge information. In a previous paper, we found the same conclusions as those of Tatler et al.'s (2005) work when examining, through a model of concurrent features, which one contributes the most to the prediction of fixation locations during free exploration of natural color scenes (Ho-Phuoc, Guyader, & Guérin-Dugué, 2010). In this work, we proposed a biologically inspired model to compute the visual saliency of static natural scenes by simulating the functions of retinal and cortical cells. According to this model, a visual stimulus is first broken down into three channels: luminance and two color-opponent channels. The retina preprocessing is applied to the luminance channel in order to enhance neuronal tuning for high spatial frequencies. Each channel is then broken down into spatial frequency and orientation maps with cortical-like filters. Finally, we combined these maps across orientations and spatial frequencies to build six different feature maps: low spatial frequency luminance, high spatial frequency luminance, low spatial frequency green–red, high spatial frequency green–red, low spatial frequency blue–yellow, and high spatial frequency blue–yellow. Then by using a statistical model (“Expectation–Maximization”) that took into account these six feature maps as well as the central fixation bias (Tatler, 2007) and a uniform distribution, we

quantified the contributions of these factors to best explain eye fixation recorded on a large panel of observers. Our study showed that color-opponent channels contribute little to eye fixations, and by contrast, high spatial frequency luminance plays a far more important role. The fact that color does not significantly contribute to the explanation of eye movements on natural scenes might be quite surprising. In fact, we live in a colorful environment and color is largely used and manipulated in movie or arts in general. That is why, to go further, we examined the influence of color on eye fixations in a free viewing experiment in which natural scene images were presented in *abnormal colors*. We would like to know whether the role of color changes when it is perceived unnaturally. The term “abnormal color” refers to the unusual appearance of color in an image, i.e., in such a way that is different from what we may see in reality. For example, in our experiment, the sky, which may often be considered blue, becomes red. This term is reused from Oliva and Schyns' (2000) study. In fact, it is shown that during scene recognition tasks, reaction time was shorter for color images than for grayscale images, and reaction time for the latter was shorter than for abnormal color images (Goffaux et al., 2005; Oliva & Schyns, 2000). However, this conclusion was limited to a specific category of images, called color-diagnostic images. Color diagnosticity refers to the degree to which an object was associated with a specific color; for example, a banana or an orange has high color diagnosticity and a lamp has low color diagnosticity (Tanaka & Presnell, 1999). Other research studied the role of color diagnosticity in object recognition (Tanaka & Presnell, 1999; Theriault, Yaxley, & Zwaan, 2009). They found that color information is important for object recognition providing a shorter reaction time and they also found that when an object is presented in an incongruous color (what we called abnormal color) the reaction time was longer than in an achromatic version; they found these results for objects with strong color associations (high color diagnostic objects).

The aim of our study is to examine whether color and abnormal color have an effect on eye movements during free-viewing scene exploration. We tested whether or not color and abnormal color modify eye movements by comparing fixation locations between grayscale scenes and the same scenes in color and abnormal color. We also analyzed the scanpaths and other eye movement parameters like fixation durations and saccade amplitudes.

Experiment

We ran an experiment where we recorded eye movements of three groups of participants freely exploring natural scene images. One group looked at the grayscale

scenes, one group looked at the same scenes in color, and another group looked at the scenes in abnormal colors.

Participants

Thirty participants (students from our university), aged from 21 to 30 years old (average: 22, standard deviation: 3), took part in the experiment. All participants had normal or corrected-to-normal visual acuity. Each participant was given written instructions at the beginning of the experiment. Participants were equally divided into three groups of 10 participants corresponding to three different stimulus conditions: grayscale scenes, color scenes, and abnormal color scenes.

Stimuli

There are 60 natural images consisting of a Kodak database (<http://www.cipr.rpi.edu/resource/stills/kodak.html>) and other personal photographs of natural scenes. Images were coded in a 24-bit colored version. Each image has a size of 1024×768 pixels (“landscape” type) or 768×1024 pixels (“portrait” type).

From the database of color scenes, we created two other databases, one consisting of grayscale scenes and one consisting of abnormal color scenes. Many physiological studies found that in the human visual system a visual stimulus is processed in three components: luminance, red–green chrominance, and blue–yellow chrominance; this separation begins at the output of the retina and remains in the visual cortex (Chatterjee & Callaway, 2003; Dacey, 1996; Dacey & Packer, 2003). In our saliency model (Ho-Phuoc, Guyader et al., 2010), we also took into account one luminance component and two chrominance ones. The coefficients used to compute the luminance component (Equation 1) come from the extraction of luminance in the NTSC system that defines the color television broadcasting system widely used in North America. For the chrominance components, the literature used several methods to represent color opponency. The red–green channel is computed as the difference between R and G channels, so is the blue–yellow channel (Itti et al., 1998; Tatler et al., 2005). In Le Meur

et al.’s (2006) saliency model, the authors also used red–green and blue–yellow channels by replacing R, G, B channels with L, M, S cones (these cones are sensitive to long, medium, and short wavelengths, respectively). Moreover, there exist several other color spaces, such as DKL, Luv, and Lab, which represent color opponency simulating the color coding done by the retina. In this paper, as in our saliency model presented in a previous study (Ho-Phuoc, Guyader et al., 2010), we adopt a simple approximation of color opponency in a similar way as Itti’s model (Itti et al., 1998; Tatler et al., 2005). Hence, color images are described in luminance and two chrominance channels according to the following equations:

$$L = 0.2989R + 0.5870G + 0.1140B,$$

$$RG = R - G, \tag{1}$$

$$BY = B - \frac{R + G}{2},$$

where R, G, B are the three plans of the normal color image in the RGB color space. Hence, a grayscale image is created by keeping luminance component *L* in Equation 1. For the abnormal color images, RG and BY are permuted. We ensured that RG and BY channels kept their initial mean energies unchanged. The three channels, L, RG, BY, were then transformed back to three components: R, G, B (RGB color space). Therefore, the three databases (grayscale, color, and abnormal color) contain the same scenes with the same luminance. Figures 1a–1c show an example of a particular scene in its grayscale, color, and abnormal color versions, respectively.

Procedure

The Eyelink II (SR Research, <http://www.sr-research.com/>) was used to record eye movements with a sampling rate of 250 Hz in the monocular Pupil-CR recording mode. For each participant, a 9-point calibration was carried out before the experiment and a drift correction was done between each image. The experimenter was in



Figure 1. Example of a scene used in the three stimulus conditions: (a) grayscale, (b) color, and (c) abnormal color version.

the experimental room to control the drift and to do a new calibration if needed (central drift larger than 0.5°).

In each stimulus condition, participants were seated 57 cm from the display ($27^\circ \times 42^\circ$ visual field) with their chins supported on a fixed bar to limit possible head movements. Images were displayed on the center of an Apple 20-inch flat panel screen (resolution of 1280×1024 pixels). We carefully measured the gamma functions of our display for the three channels (red, green, and blue) using an eye-one spectrometer (GretagMacbeth, Switzerland). The three stimulus conditions were seen using the same display and the same parameters.

Each participant viewed the 60 scenes in only one stimulus condition (grayscale, color, or abnormal color) and was told to look at the scenes freely. First, a participant had to fixate a white square on a mean gray screen for 100 ms; the location of the white square is equally distributed either in the center of the display or in the four display corners (at an angular eccentricity from the center of 21°). Hence, 20% of the trials began with central fixation and 80% with peripheral fixation. Second, a scene was displayed for 5 s. Finally, a gray screen was displayed for 1 s before the presentation of the next image.

Besides 60 images used for recording participants' fixations, we used two other images that helped familiarize a participant with the experiment and that were not used in the analysis. The order of appearance of images was random for each participant. The experiment was carried out in a darkened room and took about 10 min for each participant. In the following sections, eye movements from the three stimulus conditions were analyzed to examine the influence of color on eye movements in free viewing of natural images.

Comparison of fixation locations for the three stimulus conditions with a saliency model

Our experimental paradigm put observers in front of a screen and asked them to look freely at static images. In such situations, it was shown that bottom-up saliency models are able, to a certain extent, to predict observers' fixation locations (Itti et al., 1998; Le Meur et al., 2006; Marat et al., 2009; Parkhurst et al., 2002; Torralba et al., 2006), even other studies (Tatler, Hayhoe, Land, & Ballard, 2011) showed the limitations of this kind of models in more ecological situations. In this section, we use a saliency model guided by low-level visual features, to compare fixation locations recorded during the three stimulus conditions. Our previous study (Ho-Phuoc, Guyader et al., 2010) revealed that high spatial frequency luminance contributes the most to observers' fixation locations. Consequently, a saliency model based on this

feature is used to compare fixations in different stimulus conditions. It is important to remember that scenes in the different conditions have identical saliency maps with this model because of their identical luminance. The hypothesis is that if the addition of color in a scene changed observers' fixation locations, the prediction efficiency of a saliency model would vary from one stimulus condition to another.

Criteria to evaluate a saliency model

We compare fixations recorded for the three stimulus conditions with a common luminance-based saliency map. Criteria are needed to evaluate the correspondence between recorded fixations and a saliency map. We chose three classical criteria: *Normalized Scanpath Saliency* (NSS; Peters, Iyer, Itti, & Koch, 2005) evaluating the saliency predicted at fixation locations, *Torralba's Criterion* (TC; Torralba et al., 2006) quantifying the percentage of fixation locations predicted by a saliency model over all fixations, and score *s* (Hügli, Jost, & Ouerhani, 2007) measuring the excess of saliency found at the fixation points with respect to arbitrary points. The first and third criteria preserve the sensitivity of pixel saliency, while the second segments an image into a salient region and a non-salient one. Contrary to the first two criteria, the score *s* depends on the amplitude of a saliency map. For each scene, the three criteria were computed between the saliency map of the scene and the fixations of an observer exploring the scene in a particular stimulus condition. Furthermore, to be able to say whether the value obtained for a particular criterion is "good" or "not," we estimated the minimum and maximum values of each criterion (boundaries). For the lower boundary, criteria were computed to compare the saliency map of an image with "random fixations" obtained using fixations recorded on another image (Reinagel & Zador, 1999; Tatler et al., 2005). For the upper boundary, observers' fixation locations were compared to a saliency map created from the fixation locations of other subjects by Parzen's method (putting a Gaussian function at each fixation location). This map is called "inter-observer map" (Figure 2). In fact, because of strong consistency between observers' fixations, there is no computational saliency model that can predict an observer's fixation locations better than the model using fixations from other subjects (Torralba et al., 2006). The criteria for random fixations and inter-observer maps are computed on fixations recorded during the grayscale stimulus condition.

Results

As in the previous study (Ho-Phuoc, Guyader et al., 2010), the first eight fixations of an observer for an image in a stimulus condition were used to compute criteria.

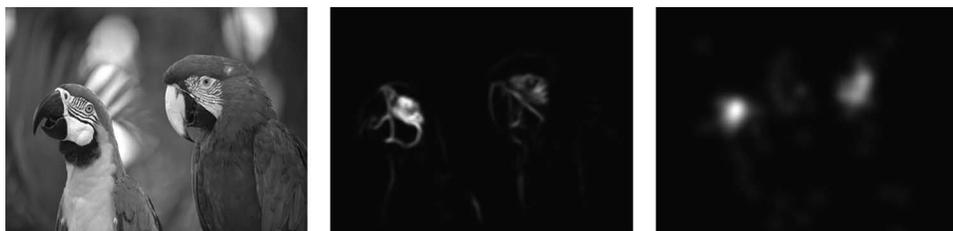


Figure 2. Examples of saliency map (b) and inter-observer map (c) for a grayscale image (a).

Usually, bottom-up visual attention models are used to predict the first fixations of scene exploration. Criteria were averaged over the 10 observers and the 60 scenes for each condition. Figure 3 illustrates the NSS, TC, and score *s* criteria for our luminance-based saliency model compared with fixations in the three stimulus conditions (G for grayscale scenes, C for color scenes, A for abnormal color scenes), for the saliency model compared with random fixations (R for random fixations), and for inter-observer maps compared with fixations in the grayscale stimulus condition (H for the human inter-observer map). For each criterion, results are always higher than chance level (0 for NSS, 20% for TC, and 0 for score *s*). Results for the three stimulus conditions are clearly between the “Random” and “Human” results, which may represent the lower and

upper criterion boundaries. This needs to be confirmed by a statistical test. We used the Kolmogorov–Smirnov test (KS test), which is a non-parametric test and is based on the empirical distribution function in order to test whether two samples are drawn from the same distribution. An advantage of the KS test is that it does not require the knowledge of *a priori* distribution. Here, the KS test showed significant differences between “Random” or “Human” condition and each of the three stimulus conditions (KS test, $p = 0$).

Figure 3 displays high values for the three criteria NSS, TC, and score *s* when comparing the fixations with the inter-observer map (Human); this confirms the consistency between observers’ fixation locations during free viewing of natural scenes. The consistency between

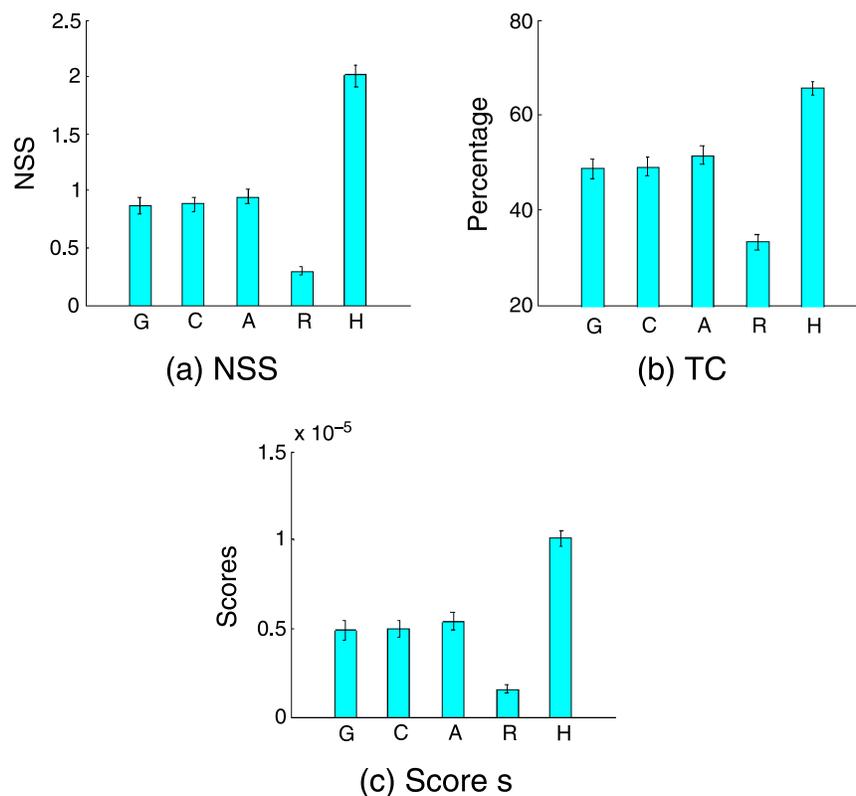


Figure 3. Three criteria, NSS, TC, and score *s*, were computed to compare fixations obtained for the three stimulus conditions (grayscale “G,” color “C,” and abnormal “A”) with a luminance-based saliency map, random fixations “R” with the saliency map, and fixations in the grayscale stimulus condition with an inter-observer map (Human “H”). These two last values correspond to the lower and upper boundaries of criteria. The error bars at 95% are computed by a “bootstrap” estimate (10,000 replications).

observers illustrates the fact that low-level visual features guide eye movements in the case of static scene exploration. This is also illustrated by the efficiency of our luminance-based saliency map to predict eye fixation locations for the grayscale stimulus condition. For all criteria, the results of our saliency model compared with fixation locations are much higher than those compared with random fixations. Nevertheless, it is interesting to observe that the “Random” result is higher than the chance value. This result can be interpreted by two biases: first, natural scenes usually consist of salient regions at the center, and second, observers have a tendency to look at the scene center whether the experiment is free viewing or task dependent (Tatler, 2007). Consequently, using fixations from another image as non-fixated locations to compare with the saliency map of an image can bring about a more reasonable value of the criterion than using artificially random fixations. For the three criteria, there is no difference between the three stimulus conditions. We observe that NSS values for the grayscale and color stimulus conditions are nearly identical (KS test, $p = 0.59$). The NSS value of the abnormal stimulus condition is somewhat higher, but the difference is not significant (KS test, $p = 0.16$ between grayscale and abnormal color conditions; KS test, $p = 0.15$ between color and abnormal color conditions). Similar results are obtained with the TC and score s criterion.

This means that the addition of color when viewing a natural scene does not change observers’ fixation locations compared to a grayscale scene. This result might be explained by the fact that people still perceive color even when looking at grayscale images (Hansen, Olkkonen, Walter, & Gegenfurtner, 2006); when viewing a forest with a sky, one might immediately perceive blue sky and green forest. However, this does not explain why the abnormal color does not change fixation locations. Perhaps, fixation locations may be mainly explained by luminance contrast, which is identical in the three conditions. In the present experiment, observers were asked to look freely at natural scenes without any particular task and without paying particular attention to color information. Moreover, the scenes chosen represent natural environments and might not have high color diagnosticity. This might be one of the reasons to explain why we do not observe significant differences between the fixations recorded during grayscale, color, and abnormal color stimulus conditions.

All these first results might be verified using only the observers’ fixations without taking into account a saliency model. The next section is dedicated to the inter-observer consistency criterion (Tatler et al., 2005) based on fixation locations of all the observers on the different scenes.

Inter-observer consistency

Several studies have observed that at the beginning of viewing, observers are likely to look at the same areas;

this trend decreases as time goes on. This can be described by the inter-observer consistency criterion that is computed as follows: The metric is presented as a function of fixation rank (i.e., time) and is computed in a similar way to inter-observer maps (see [Criteria to evaluate a saliency model](#) section). At each fixation rank k , fixations of an observer i are collected from all scenes (that this observer viewed). From this fixation set of observer i , we create a density map for this observer using Parzen’s method. This density map also has the image size (1024×768) and is normalized in such a way that its sum is equal to one. Then, the fixations of all observers other than observer i are grouped together and a density map for this fixation set is computed in the same way as above. Thus, there are two density maps: the first one (P_i) for fixations of observer i and the second (P_i^*) for fixations of all observers other than subject i . The Kullback–Leibler (KL) divergence (Kullback & Leibler, 1951) measures the distance between these two density maps:

$$\begin{aligned} D(k, s_i) &= D_{KL}(P_i, P_i^*), \\ &= \frac{1}{2} \left(\sum P_i \log \frac{P_i}{P_i^*} + \sum P_i^* \log \frac{P_i^*}{P_i} \right), \end{aligned} \quad (2)$$

where $D(k, s_i)$ is the KL divergence for observer i at fixation rank k . The KL divergence at fixation k is the average of the divergences of all observers at this fixation:

$$D(k) = \frac{1}{N_S} \sum_{i=1}^{N_S} D(k, s_i), \quad (3)$$

where N_S is the number of observers.

The higher the KL divergence is, the lower the inter-observer consistency is. [Figure 4](#) represents the inter-observer consistency curves in the three conditions with the first eight fixations.

Results confirm the fact that inter-observer consistency decreases during scene exploration (increase of the Kullback–Leibler divergence; [Figure 4](#)). When comparing the three curves, we observe that the curves in the three stimulus conditions are very similar (KS test, $p > 0.05$). Here, we obtained the same pattern of results as above: the addition of color, even abnormal color, does not influence fixation locations or inter-observer consistency.

The location of fixation is an aspect of visual exploration, but it is not the only one. It might be interesting to test whether color and abnormal color modify more generally the scanpaths (the fixation sequence including the temporal aspect, i.e., the fixation rank), the fixation durations, and the saccade amplitudes. However, most of these quantities are barely predicted by a saliency model; for instance, there is no correlation between computed saliency and fixation duration (Itti, 2005). In the following

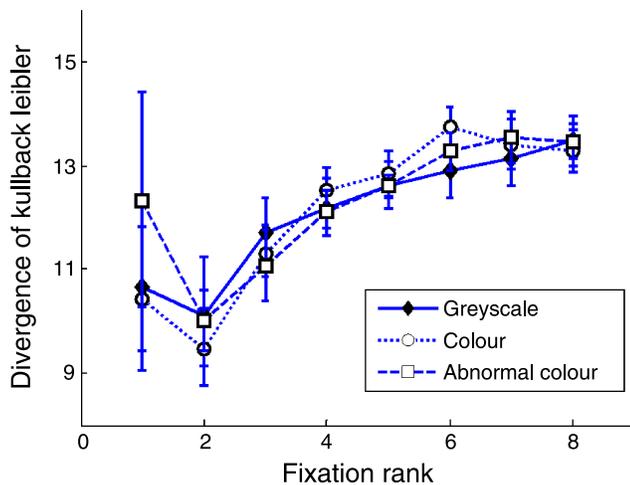


Figure 4. Inter-observer divergence as a function of fixation rank for the three stimulus conditions (grayscale, color, and abnormal color images). The error bars at 95% are computed by a “bootstrap” estimate (10,000 replications).

section, we test whether color influences other eye movement parameters, and this does not require saliency modeling.

Statistics of eye movement parameters

Fixation locations are very often used to study eye movements or to evaluate the prediction quality of a saliency model. In the previous section, it was shown that there is no difference in fixation locations between the three stimulus conditions: grayscale, color, and abnormal color. However, fixation locations are not enough to characterize eye movements. In the literature, the role of color and abnormal color was disclosed with different criteria, e.g., time reaction in scene recognition or performance in categorization (Goffaux et al., 2005; Oliva & Schyns, 2000). As there is no such criteria in our experiment of free viewing, we looked at several properties other than fixation locations in order to compare eye movements in the three stimulus conditions. First, we looked at the scanpaths using a specific metric to compute distances between scanpaths (“ScanMatch” from Cristino, Mathôt, Theeuwes, & Gilchrist, 2010) and also the inertia of scanpaths that gives information about the extent of scene exploration as a function of the stimulus condition. Second, we analyzed other eye movement properties based on fixations and saccades. Using the EyeLink II (SR Research), we can extract, together with fixation locations used in the previous section, fixation durations and saccade amplitudes.

Scanpaths

If the color layout seems to have no influence on fixation positions, we are interested in the temporal

sequence of the fixations. Hence, we wanted to see if color information influenced the global scanpath recorded for a scene. The similarity between scanpaths was assessed from the “string edit distance” using the “ScanMatch” toolbox (Cristino et al., 2010). In information theory, the “edit distance” between two strings of characters is the number of operations required to transform one of them into the other. This distance measurement was applied to compute the similarity between scanpaths in eye movement research (Cristino et al., 2010). For each scene and under the same stimulus condition, we computed the similarity between the scanpaths of all pairs of observers. We obtained three distributions that correspond to the within-condition similarity values because it was computed for the same stimulus condition. Similarly, we repeated this computation process for each scene but between two scanpaths recorded for two different stimulus conditions. We also obtained three distributions that correspond to between-condition similarity values. For each scene, the aim was to test whether the between-condition similarities were lower than the within-condition similarities. Finally, we found that, for each scene, the differences between the within-condition and between-condition similarities were not significant; this was due to a large inter-individual variation of the scanpaths.

We also analyzed the inertia of each scanpath in order to evaluate the overall spatial distribution of fixations under the different stimulus conditions. For each scene, in each stimulus condition, the mean inertia over all fixations of observers was computed, that is, the average of all the square distances of fixation locations from the mean fixation location recorded for a scene. We obtained a mean inertia for the abnormal color scenes smaller than that for color scenes, which was then smaller than for grayscale scenes (Figure 5); however, these differences were not statistically significant.

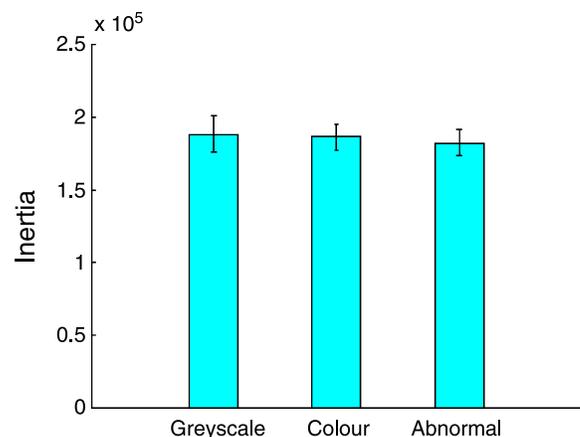


Figure 5. Mean inertia of fixations for the three stimulus conditions (grayscale, color, and abnormal color scenes). The error bars at 95% are computed by a “bootstrap” estimate (10,000 replications).

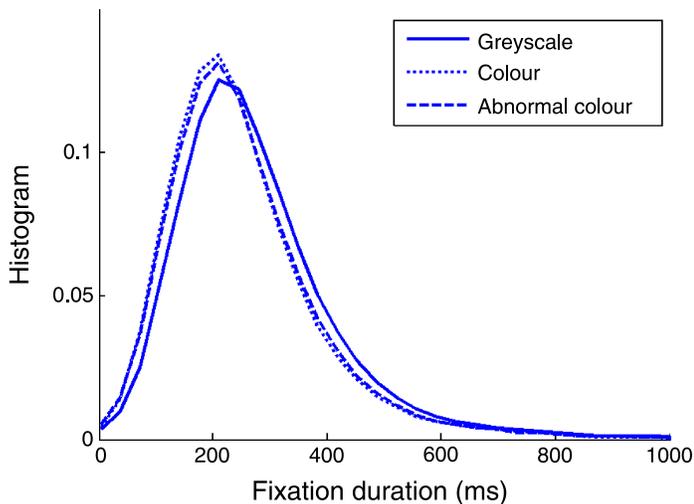


Figure 6. Histogram of fixation durations for the three stimulus conditions (grayscale, color, and abnormal color scenes).

Color and abnormal color information does not influence the scanpaths of observers (i.e., the temporal sequences of fixations). Moreover, it does not modify the overall spatial distributions of fixations when exploring natural scenes. To go further, we analyzed other eye movement parameters.

Fixation duration

Fixation durations are extracted from 10 subjects looking at 60 scenes during 5 s for each stimulus condition. On average, we had 8201, 9117, and 9029 fixations, respectively, for the grayscale, color, and abnormal stimulus conditions.

Figure 6 displays a histogram of the fixation durations for the three stimulus conditions (histograms are smoothed by convolution with a Gaussian function; this is for graphic representation and does not affect the distribution of fixation durations). From these curves, we observe that in all three conditions the most frequent fixation duration is about 200 ms; few fixations are shorter than 100 ms or longer than 400 ms. These results are similar to those reported in the literature (Andrews & Coppola, 1999; Tatler & Vincent, 2008). In order to examine the difference between the three distributions of fixation durations, we used the Kolmogorov–Smirnov test (KS test). Results show that there exists a difference between the distributions of fixation durations in grayscale and color conditions (KS test, $p = 1.30e-36$) and between grayscale and abnormal color conditions (KS test, $p = 2.74e-26$). By contrast, we did not find any significant difference between fixation durations of color and abnormal color stimulus conditions (KS test, $p = 0.04$). On average, fixation durations are smaller for color and abnormal color stimulus conditions than for grayscale stimulus condition.

In order to consider the temporal variation of fixation durations, we computed the mean fixation duration according to fixation order (Figure 7). Rather than taking into account all the different fixations, we grouped consecutive fixations; hence, the mean fixation duration was computed for the first three fixations of the scene exploration, then for the next three fixations, and so on, until the fifteenth fixation. Figure 7 shows a similar shape for the three temporal distributions of fixation duration: It increases at first and then decreases as time goes by. Furthermore, as in Figure 6, Figure 7 shows that observers make longer fixations, and therefore fewer fixations, in the grayscale stimulus condition. It is interesting to note that at the very beginning of scene exploration, there is a significant difference between the mean fixation duration of the color condition and the two other stimulus conditions. Observers made shorter fixation duration for color scenes (KS test: $p = 6.46e-19$ between the grayscale and color conditions; $p = 5.43e-5$ between the color and abnormal color conditions; and $p = 2.75e-6$ between the grayscale and abnormal color conditions). Then, for the next three fixations, the difference appeared between the grayscale scenes and the other two stimulus conditions. Fixation durations were longer for grayscale scenes than for color and abnormal color scenes (KS test: $p = 9.80e-10$ between the grayscale and color conditions; $p = 4.56e-9$ between the grayscale and abnormal color conditions; and $p = 0.88$ between the color and abnormal color conditions). Until the thirteenth fixation, the fixations are longer for grayscale scenes compared to color and abnormal color scenes. Finally, for the last fixations

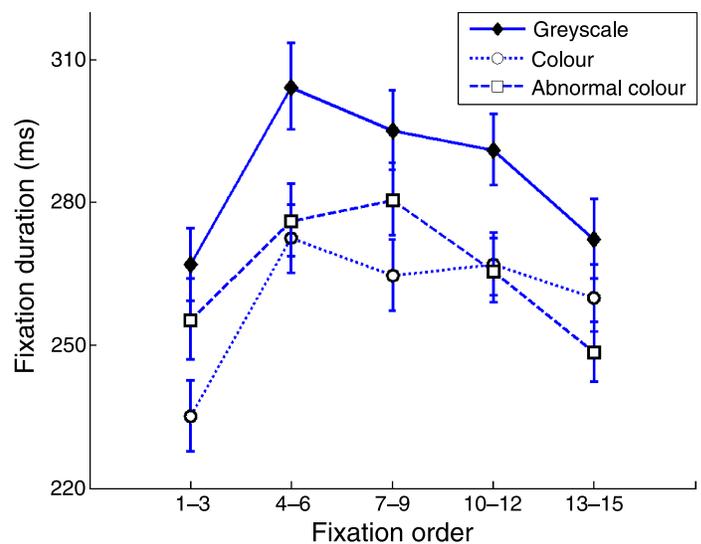


Figure 7. Fixation duration according to fixation order (we only kept the first fifteen fixations of each scene and we split these fixations into five classes: the first three fixations, the next three, and so on) for the three stimulus conditions (grayscale, color, and abnormal color images). The error bars at 95% are computed by a “bootstrap” estimate (10,000 replications).

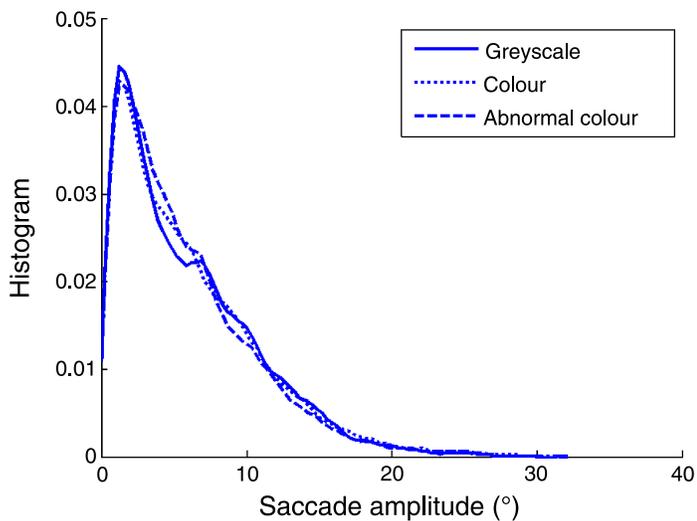


Figure 8. Histogram of saccade amplitudes for the three stimulus conditions (grayscale, color, and abnormal color scenes).

there is no longer any difference in fixation duration between the three stimulus conditions.

To summarize, color information modified the fixation durations. Fixations were shorter for normal color scenes at the very beginning of scene exploration (averaged on the first three fixations), then fixations were shorter for abnormal and normal color scenes compared to grayscale scenes (until the thirteenth fixations, i.e., around 3 to 3.5 s). When exploration ended, we could not show any significant difference in fixation durations between the different stimulus conditions.

Saccade amplitude

Another property of eye movements that is often examined is the distribution of saccade amplitude. Saccade amplitude is the distance (in angular degrees) between two successive fixation locations.

As for fixation durations, saccade amplitudes are also gathered from all images and all observers for each stimulus condition. The smoothed distributions of saccade amplitudes in the three conditions are shown in Figure 8. Each histogram has a long-tailed distribution that has been reported in the literature: Most saccades have amplitudes smaller than 15° (Bahill, Adler, & Stark, 1975). We notice that the experimental conditions (e.g., visual field) do not result in small saccade amplitudes since the image size may be as large as 34° . These distributions of saccade amplitudes may also be simulated by a Gamma distribution (Ho-Phuoc, Guérin-Dugué, & Guyader, 2010). By using KS tests (the sample size was 8709, 9618, and 9547, respectively, for the grayscale, color and abnormal stimulus conditions), we observe the difference in distributions of saccade amplitudes between grayscale and abnormal color stimulus

conditions and between color and abnormal color stimulus conditions (KS test, $p = 1.51e-4$ and $p = 3.05e-4$, respectively). By contrast, there was no difference between grayscale and color stimulus conditions (KS test, $p = 0.15$).

Relation of saccade amplitude and fixation duration

Several studies showed close relations between fixations and saccades (Tatler & Vincent, 2008; Velichkovsky, Joos, Helmert, & Pannasch, 2005). Thus, it is interesting to use a combination of these two properties to compare eye movements in the three stimulus conditions. Here, we focus on the relation between current fixation duration and following amplitude saccade; this relation is very often examined in the literature. Figure 9 presents the following saccade amplitude as a function of current fixation duration. It is shown that in the grayscale and color stimulus conditions, saccades have small amplitudes when they follow fixations with too short or too long duration. Most saccades with large amplitude are preceded by fixations with a duration of between 80 and 150 ms. These results are consistent with the literature (Tatler & Vincent, 2008; Velichkovsky et al., 2005). However, the “saccade amplitude–fixation duration” curve in the abnormal color stimulus condition seems to be slightly different. While it is very similar to the curve of the grayscale or color stimulus condition for fixations that are longer than 100 ms, it differs from others for fixations shorter than 100 ms: Very short fixations are followed by long saccades. Perhaps,

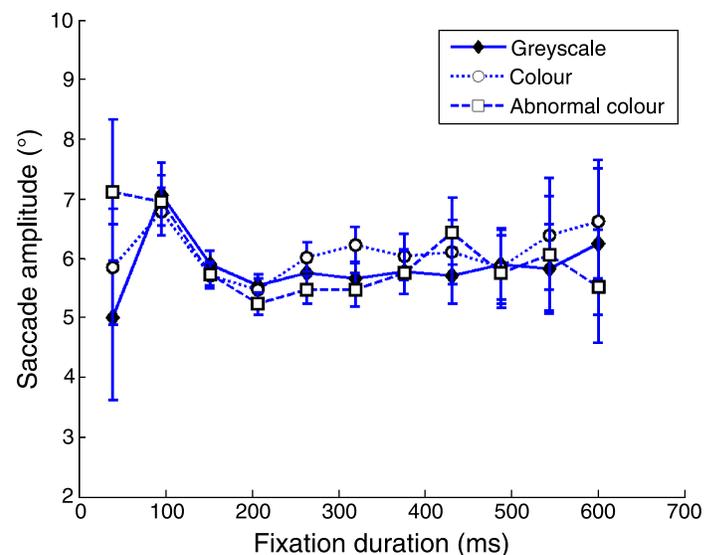


Figure 9. Following saccade amplitude as a function of current fixation duration for the three stimulus conditions (grayscale, color, and abnormal color scenes). The error bars at 95% are computed by a “bootstrap” estimate (10,000 replications).

this reflects an impact of abnormal color on eye movements. We would also point out that it is necessary to focus mainly on fixations with duration shorter than 400 ms because few fixation durations are beyond this value (Figure 6).

Discussion

In this paper, we examined several properties of fixations and saccades in order to compare eye movements in the three stimulus conditions of our experiment: grayscale, color, and abnormal color images. Scenes had the same luminance in these three conditions. By adding color or modifying color abnormally, we studied the influence of color on eye movements during free exploration of natural scenes.

In the [Comparison of fixation locations for the three stimulus conditions with a saliency model](#) section, we used a saliency model to compare the observers' fixation locations in the three stimulus conditions. Results showed that addition of normal and abnormal color did not have an impact on fixation locations. Moreover, the overall spatial distribution of fixations was not impacted by the addition of color. These first experimental results, about the role of color in explaining eye movement during free natural scenes, are consistent with our previous study in which we quantified the contributions of several low-level visual features to saliency at fixation locations and found that color does not much explain fixation selection compared to luminance (Ho-Phuoc, Guyader et al., 2010). A similar result was previously revealed by Tatler et al. (2005). In other words, addition or removal of color in a visual scene does not influence eye fixation locations when visually exploring the scene. While color does not modify fixation locations when looking at a scene, it may not be necessary to take into account color in a saliency model, particularly when it is used only to predict the fixation locations of observers. This is also an answer to a larger question that has been discussed: Is it mandatory to combine all features in order to build a saliency map? Here, our results are in line with the previous ones of Baddeley and Tatler (2006) and Tatler and Vincent (2009) who argued that a model of edges alone is at least as good as a full saliency model that fuzzes several elementary feature maps into a master saliency map. In consequence, salient regions can be almost determined by luminance, and this considerably reduces the number of features computed in a saliency model, which in turn simplifies its computational complexity.

The influence of color seems to be related to characteristics other than fixation location. It is interesting to notice that, in our experiment, the addition of color, no matter whether it is normal or abnormal, decreases fixation

duration. For the normal color condition, the decrease of fixation duration might be explained by the “surface-plus-edge-based” theory (for a review, see Tanaka et al., 2001). In fact, according to this theory, objects are recognized faster when presented in color than when presented in grayscale. In this case, the object representation contains information not only about the object “shape” but also about the surface properties like color. For the “Shape + Surface” model of object recognition, color provides one of the perceptual inputs into the object representation system. The extra information from color helps subjects to recognize objects. However, this model maintains that color plays a supporting role in the recognition of high color-diagnostic objects and scenes. This explained why in different studies they found that abnormal color might impair object or scene recognition. For example, Oliva and Schyns (2000) showed that scenes that are rich in color-diagnostic content are best recognized in their normal colors than in abnormal color. The same was found for object recognition by Tanaka and Presnell (1999). Hence, object recognition is jointly determined by the bottom-up influence of perceptual color and the top-down influence of color knowledge. Following this, the “Shape + Surface” theory does not justify why in our experiment we observed the same fixation duration decrease for normal and abnormal colors. However, it is important to remind that in our experiment observers were not given any particular task. They explored the scene without the aim to categorize the scene or to recognize an object. Hence, it is difficult to link our results with previous ones.

Moreover, in this paper, we found that the abnormal color condition makes the saccade amplitude shorter than in the two other conditions. This might be explained as follows: When viewing an abnormal color object, observers are likely to try to search several regions that are around and not far from the object. This strategy might help one obtain relevant information that is missed in the current object due to abnormal color. Using this hypothesis, we can explain the shorter fixation duration in the abnormal color condition above: As soon as abnormal information is detected, the human eye rapidly moves to neighboring regions to look for more relevant information, which can, in turn, result in such shorter fixation duration.

Until now, saliency models have often been compared to fixation locations, but they have revealed no ability to predict fixation durations or saccades. On the other hand, it seems possible to predict fixation location using saccade information (Tatler & Vincent, 2009). Indeed, in their study, Tatler and Vincent showed that a model based only on information about the amplitude and direction of saccades, and therefore blind to current visual information, outperformed popular saliency-based approaches. Consequently, in order to improve the predictive capacity, it is necessary for a saliency model to use not only the

visual feature information (mainly the luminance information) and to combine with saccade information. Fixation durations and saccade amplitudes, as well as their relation, have been studied in the literature (Pannasch, Helmert, Roth, Herbold, & Walter, 2008; Tatler & Vincent, 2008; Velichkovsky et al., 2005). The current paper showed similarity in tendency of eye movements between the grayscale and color stimulus conditions through the relation between fixation duration and the amplitude of the following saccade; both conditions provided curves representing this relation close to one reported by Velichkovsky et al. (2005) or Tatler and Vincent (2008). By contrast, the abnormal color stimulus condition disclosed a difference: Very short fixation duration was followed by large saccade amplitudes. In fact, the relation between fixation duration and saccade amplitude can reveal two principal types of periods in eye movements: periods of local scanning (characterized by long fixation duration and short saccade amplitude) and periods of relocation to new locations in a scene (characterized by large saccade amplitude; Tatler & Vincent, 2008; Velichkovsky et al., 2005). Hence, the difference reported with the abnormal color condition in the present paper might be related to a different viewing behavior when observers explore abnormal color scenes. It will be interesting to test this hypothesis by considering more in detail the relation of fixation duration and saccade amplitude, for example, taking into account two consecutive fixations or saccades (Tatler & Vincent, 2008). Looking for insight into such situations might contribute to a better understanding of eye movements.

Acknowledgments

This research was supported by the French Ministry of Higher Education and Research. The authors would like to thank Gelu Ionescu for the software he developed to run the experiment and also would like to thank the two reviewers for their helpful comments to improve the manuscript.

Commercial relationships: none.

Corresponding author: Nathalie Guyader.

Email: nathalie.guyader@gipsa-lab.grenoble-inp.fr.

Address: Gipsa-lab, Domaine Universitaire, 961 rue de la Houille Blanche, Grenoble 38400, France.

References

- Andrews, T. J., & Coppola, D. M. (1999). Idiosyncratic characteristics of saccadic eye movements when viewing different visual environments. *Vision Research*, *39*, 2947–2953.
- Baddeley, R. J., & Tatler, B. W. (2006). High frequency edges (but not contrast) predict where we fixate: A Bayesian system identification analysis. *Vision Research*, *46*, 2824–2833.
- Bahill, A. T., Adler, D., & Stark, L. (1975). Most naturally occurring human saccades have magnitudes of 15 degrees or less. *Investigative Ophthalmology & Visual Science*, *14*, 468–469.
- Buswell, G. T. (1935). *How people look at pictures: A study of the psychology of perception in art*. Chicago: University of Chicago Press.
- Castelano, M. S., Mack, M. L., & Henderson, J. M. (2009). Viewing task influences eye movement control during active scene perception. *Journal of Vision*, *9*(3):6, 1–15, <http://www.journalofvision.org/content/9/3/6>, doi:10.1167/9.3.6. [PubMed] [Article]
- Chatterjee, S., & Callaway, E. M. (2003). Parallel color-opponent pathways to primary visual cortex. *Nature*, *426*, 668–671.
- Cristino, F., Mathôt, S., Theeuwes, J., & Gilchrist, I. D. (2010). ScanMatch: A novel method for comparing fixation sequences. *Behaviour Research Methods*, *42*, 692–700.
- Dacey, D. M. (1996). Circuitry for color coding in the primate retina. *Proceeding of the National Academy of Sciences of the United States of America*, *93*, 582–588.
- Dacey, D. M., & Packer, O. S. (2003). Color coding in the primate retina: Diverse cell types and cone-specific circuitry. *Current Opinion Neurobiology*, *13*, 421–427.
- 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://www.journalofvision.org/content/8/14/6>, doi:10.1167/8.14.6. [PubMed] [Article]
- Goffaux, V., Jacques, C., Mouraux, A., Oliva, A., Schyns, P. G., & Rossion, B. (2005). Diagnostic colors contribute to the early stages of scene categorization: Behavioural and neurophysiological evidence. *Visual Cognition*, *12*, 878–892.
- Hansen, T., Olkkonen, M., Walter, S., & Gegenfurtner, K. R. (2006). Memory modulates color appearance. *Nature Neuroscience*, *9*, 1367–1368.
- Henderson, J. M., & Hollingworth, A. (1999). High-level scene perception. *Annual Review of Psychology*, *50*, 243–271.
- Ho-Phuoc, T., Guérin-Dugué, A., & Guyader, N. (2010). A biologically-inspired visual saliency model to test different strategies of saccade programming. *Communications in Computer and Information Science (CCIS)*, *52*, 187–199.

- Ho-Phuoc, T., Guyader, N., & Guerin-Dugue, A. (2010). A functional and statistical bottom-up saliency model to reveal the relative contributions of low-level visual guiding factors. *Cognitive Computation*, *2*, 344–359.
- Hubel, D. H., Wiesel, T. N., & Stryker, M. P. (1977). Orientation columns in macaque monkey visual cortex demonstrated by the 2-deoxyglucose autoradiographic technique. *Nature*, *269*, 328–330.
- Hügli, H., Jost, T., & Ouerhani, N. (2005). Model performance for visual attention in real 3D color scenes. In J. Mira & J. R. Alvarez (Eds.), *Lecture notes in computer science* (vol. 3562, pp. 469–478). Berlin: Springer.
- Itti, L. (2005). Quantifying the contribution of low-level saliency to human eye movements in dynamic scenes. *Visual Cognition*, *12*, 1093–1123.
- 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.
- Jost, T., Ouerhani, N., Wartburg, R. V., Müri, R., & Hügli, H. (2005). Assessing the contribution of color in visual attention. *Computer Vision and Image Understanding*, *100*, 107–123.
- Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: Towards the underlying neural circuitry. *Human Neurobiology*, *4*, 219–227.
- Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *Annals of Mathematical Statistics*, *22*, 79–86.
- Le Meur, O., Le Callet, P., Barba, D., & Thoreau, D. A. (2006). Coherent computational approach to model bottom-up visual attention. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *28*, 802–817.
- Mannan, S. K., Ruddock, K. H., & Wooding, D. S. (1997). Fixation sequences made during visual examination of briefly presented 2d images. *Spatial Vision*, *11*, 157–178.
- Marat, S., Ho-Phuoc, T., Granjon, L., Guyader, N., Pellerin, D., & Guerin-Dugue, A. (2009). A modeling spatio-temporal saliency to predict gaze direction for short videos. *International Journal of Computer Vision*, *82*, 231–243.
- Oliva, A., & Schyns, P. G. (2000). Diagnostic colors mediate scene recognition. *Cognitive Psychology*, *41*, 176–210.
- Pannasch, S., Helmert, J. R., Roth, K., Herbold, A. K., & Walter, H. (2008). Visual fixation durations and saccade amplitudes: Shifting relationship in a variety of conditions. *Journal of Eye Movement Research*, *2*, 1–19.
- Parkhurst, D. J., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision Research*, *42*, 107–123.
- Peters, R. J., & Itti, L. (2008). Applying computational tools to predict gaze direction in interactive visual environments. *ACM Transactions on Applied Perception*, *5*, 8.
- Peters, R. J., Iyer, A., Itti, L., & Koch, C. (2005). Components of bottom-up gaze allocation in natural images. *Vision Research*, *45*, 2397–2416.
- Privitera, C. M., & Stark, L. W. (2000). Algorithms for defining visual regions-of-interest: Comparison with eye fixations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *22*, 970–982.
- Reinagel, P., & Zador, A. M. (1999). Natural scene statistics at the centre of gaze. *Network*, *10*, 341–350.
- Tanaka, J. W., & Presnell, L. M. (1999). Color diagnosticity in object recognition. *Perception & Psychophysics*, *61*, 1140–1153.
- Tanaka, J. W., Weiskopf, D., & Williams, P. (2001). The role of color in high-level vision. *Trends in Cognitive Sciences*, *5*, 211–215.
- Tatler, B. W. (2007). The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions. *Journal of Vision*, *7*(14):4, 1–17, <http://www.journalofvision.org/content/7/14/4>, doi:10.1167/7.14.4. [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.
- Tatler, B. W., Hayhoe, M. M., Land, M. F., & Ballard, D. H. (2011). Eye guidance in natural vision: Reinterpreting salience. *Journal of Vision*, *11*(5):5, 1–23, <http://www.journalofvision.org/content/11/5/5>, doi:10.1167/11.5.5. [PubMed] [Article]
- Tatler, B. W., & Vincent, B. T. (2008). Systematic tendencies in scene viewing. *Journal of Eye Movement Research*, *2*, 1–18.
- Tatler, B. W., & Vincent, B. T. (2009). The prominence of behavioural biases in eye guidance. *Visual Cognition*, *17*, 1029–1054.
- Therriault, D. J., Yaxley, R. H., & Zwaan, R. A. (2009). The role of color diagnosticity in object recognition and representation. *Cognitive Processing*, *10*, 335–342.

- Torralba, A., Oliva, A., Castelhana, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. *Psychological Review*, *113*, 766–786.
- Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*, 97–136.
- Velichkovsky, B. M., Joos, M., Helmert, J. R., & Pannasch, S. (2005). Two visual systems and their eye movements: Evidence from static and dynamic scene perception. *Proceedings of the XXVII Conference of the Cognitive Science Society*, 2283–2288.
- Wolfe, J. M., & Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience*, *5*, 1–7.
- Yarbus, I. A. (1967). *Eye movements and vision*. New York: Plenum Press.