

## **Overt visual attention under natural conditions**

### **Dissertation**

zur Erlangung des Grades eines Doktors der Kognitionswissenschaft  
eingereicht am Fachbereich Humanwissenschaften  
der Universität Osnabrück

vorgelegt von

**Hans-Peter Frey**

Osnabrück, November 2008

## Abstract

In a natural environment, humans are not able to process all information available to the visual system simultaneously. As a consequence, we attend to different subparts of the input one after the other. Under natural viewing conditions, these shifts of attention are associated with changes in fixation. This so-called “overt” attention therefore provides an objective measure of attention shifts.

In my thesis, I investigate the influence of color on overt visual attention. I present human subjects with different categories of color-calibrated images and record their eye-movements. If an image feature like red-green color contrast has an influence on attention i.e. is salient, then it should be strongly elevated at fixated locations. However, since this measure is only correlational, I present the images in several different levels of modifications. The modifications, systematic changes across the entire image, influence only the global appearance of the stimuli, while leaving the local features under investigation unaffected. In the first series of experiments, images of 7 different categories (Face, Flower, Forest, Fractal, Landscape, Man-Made object, and Rainforest) are presented either in natural color or grayscale. With regard to the influence of color on overt attention, I find two extreme categories: while in Rainforest images all color features examined (red-green and blue-yellow color contrast, as well as saturation) are salient, none is salient in Fractal. 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

In the second series of experiments, I use only Rainforest images. These are presented to color-normal or deuteranope subjects with several modifications in the color domain. I find a causal influence of color-contrast on overt attention, processed in a way that it is not specific to either the red-green or blue-yellow color channel. In the case of color blindness, a slower high-level compensatory mechanism affects the selection of fixation points.

These experiments show that there is no single color feature, which influences overt attention in all possible environments. I provide evidence that different levels of the visual hierarchy are involved in the computation of saliency. The experiments conducted also highlight the importance of higher-level mechanisms.

## Table of contents

List of Figures & List of Tables .....	5
1. General Introduction .....	6
1.1 Properties of Visual Attention .....	6
1.1.1 Spatial Attention .....	6
1.1.2 Object-based attention .....	7
1.1.3 Feature-based attention .....	8
1.2 Overt and Covert Visual Attention .....	8
1.2.1 Covert visual attention .....	8
1.2.2 Overt visual attention .....	8
1.3 Bottom-Up and Top-Down control of attention .....	9
1.3.1 Bottom-up .....	9
1.3.2 Top-down .....	10
1.4 Color and attention .....	11
1.5 Aim of this thesis .....	12
2. What's color got to do with it? The influence of color on visual attention in different categories .....	13
2.0 Abstract .....	13
2.1 Introduction .....	13
2.2 Methods .....	17
2.2.1 Subjects .....	17
2.2.2 Color representation .....	17
2.2.3 Stimuli .....	17
2.2.4 Eye-tracking .....	18
2.2.5 Experimental Design .....	20
2.2.6 Definition of features .....	20
2.2.7 Feature analysis .....	21
2.2.8 Congruency of fixation locations between conditions and observers .....	22
2.2.9 Saliency map .....	23
2.2.10 Control experiment .....	24
2.3 Results .....	24
2.3.1 Eye-movements of human subjects .....	24
2.3.2 Stimulus Features .....	27

2.3.2.1 Luminance features .....	27
2.3.2.2 Color features .....	29
2.3.3 Saliency map model .....	31
2.3.4 Summary: The influence of color on man and model.....	34
2.4 Discussion .....	35
2.4.1 The influence of color in different categories.....	35
2.4.2 Saliency map.....	37
2.4.3 Is Rainforest special? .....	38
2.4.4 Color space and experimental task.....	38
2.4.5 Color features.....	39
2.5 References .....	40
3. A hue invariant causal influence of color on eye movements and a high level compensatory mechanism while studying images of rainforest. ....	44
3.0 Abstract .....	44
3.1 Introduction .....	44
3.1.1 Visual Attention.....	45
3.1.2 Color processing .....	46
3.1.3 Color and attention.....	47
3.1.4 This study .....	48
3.2 Experiment 1 .....	48
3.2.1 Introduction.....	48
3.2.2 Methods .....	48
3.2.2.1 Color representation .....	48
3.2.2.2 Stimuli .....	49
3.2.2.3 Subjects.....	50
3.2.2.4 Experimental Design .....	50
3.2.2.5 Eye-tracking.....	51
3.2.2.6 Definition of features.....	51
3.2.2.7 Feature analysis.....	52
3.2.2.8 Congruency of fixation locations between conditions and observers .....	53
3.2.3 Results .....	54
3.2.3.1 Analysis of image features.....	54
3.2.3.2 Analysis of fixation distributions .....	56
3.2.4 Discussion.....	57

3.3 Experiment 2 .....	58
3.3.1 Introduction.....	58
3.3.2 Methods .....	58
3.3.2.1 Stimuli .....	58
3.3.2.2 Experimental Design .....	59
3.3.2.3 Image statistics .....	59
3.3.3 Results .....	60
3.3.3.1 Analysis of image features.....	60
3.3.3.1.1 Hue.....	60
3.3.3.1.2 Color contrast.....	62
3.3.3.1.3 Asymmetry.....	64
3.3.3.1.4 Saturation .....	65
3.3.3.2 Analysis of fixation distributions .....	66
3.3.4 Discussion.....	66
3.4 Experiment 3 .....	68
3.4.1 Introduction.....	68
3.4.2 Methods .....	68
3.4.2.1 Subjects.....	68
3.4.2.2 Stimuli .....	68
3.4.2.3 Experimental Design .....	69
3.4.3 Results .....	69
3.4.3.1 Analysis of image features.....	69
3.4.3.2 Fixations on Fruit .....	70
3.4.3.3 Analysis of fixation distributions .....	73
3.4.4 Discussion.....	74
3.5 Discussion of Experiments 1-3.....	75
3.5.1 Different levels of attentional control .....	75
3.5.2 Other image features .....	76
3.5.3 Timing of bottom-up and top-down.....	77
3.5.4 Deuteranope subjects.....	78
3.5.5 Rainforest images.....	79
3.6 References .....	80
4. General Discussion .....	84
5. General References.....	85

6. Acknowledgements.....	89
--------------------------	----

## List of Figures

Figure 2.1 Example images.....	19
Figure 2.2 Feature analysis. ....	22
Figure 2.3 Calculation of congruency between observers.....	23
Figure 2.4 Similarity of fixation locations in colored and grayscale conditions. ....	25
Figure 2.5 Inter-observer congruency. ....	27
Figure 2.6 Luminance features.....	28
Figure 2.7 Color features. ....	30
Figure 2.8 Saliency map model.....	33
Figure 2.9 Influence of color on humans and saliency map. ....	35
Figure 3.1 Experimental conditions in Experiment 1.....	50
Figure 3.2 Feature analysis. ....	53
Figure 3.3 Calculation of congruency between observers.....	54
Figure 3.4 Color features. ....	55
Figure 3.5 Congruency between fixation locations of color-normal subjects. ....	57
Figure 3.6 Experimental conditions in Experiment 2.....	59
Figure 3.7 Influence of hue on overt attention.....	61
Figure 3.8 Influence of hue on overt attention.....	62
Figure 3.9 Influence of color-contrasts on overt attention.....	63
Figure 3.10 Peakiness of color contrasts. ....	64
Figure 3.11 Analysis of the asymmetry effect. ....	65
Figure 3.12 Influence of saturation on overt attention. ....	66
Figure 3.13 Color features in deuteranope observers.....	70
Figure 3.14 Color features at each fixation.....	71
Figure 3.15 R-G contrast at each fixation on a fruit object. ....	72
Figure 3.16 Congruency between fixation locations in deuteranope observers.....	74

## List of Tables

Table 2.1 Influence of color on the saliency map model.....	32
---	----

# 1. General Introduction

## ***1.1 Properties of Visual Attention***

Attention is a common term in our everyday language. However, attention describes a variety of different phenomena like alerting, orienting, or executive control (Posner and Petersen, 1990). Attentional selection acts on different “targets”, which can either be stimulus features (e.g. color), whole objects, or spatial locations. In recent years knowledge about these different mechanisms has gathered at an enormous pace. The influence of visual attention on perception has been studied in psychophysical and electrophysiological experiments in human subjects and animals. In the next sections I am going to describe some of the major findings, which are relevant for my thesis.

### **1.1.1 Spatial Attention**

The property of attention, which has been studied in greatest detail, is spatial attention. A compelling demonstration of spatial aspects of visual attention is change blindness. When presented with movies of a static scene, in which one area or object changes slowly and gradually, most observers do not detect the change. They have the impression that the whole scene was stable all the time. Only when they spatially attend to the changing area while it is being changed observers are able to detect the image alteration. There are several other types of change blindness paradigms, which produce the same basic result: even though our visual system makes us believe that we perceive all of our visual field all the time, we only perceive what is highlighted by visual attention.

Numerous psychophysical studies have quantified the effects of spatial attention on human perception. Spatial attention studies typically have subjects attend to a subset of the spatial array, allowing for selective report of information at the focus of attention in comparison to non-attended locations (Averbach & Coriell, 1961; Eriksen & Hoffman, 1973; Sperling, 1960). The subpart the spatial array the subject has to attend to is usually indicated by a cue, e.g. an arrow or a light flash. Attending to a spatial location improves several measurable responses. Human subjects are usually faster to respond to stimuli presented at a cued location (e.g. Posner, Snyder, & Davidson, 1980). This reaction time advantage is often used as a behavioral metric in electrophysiological studies. In addition, pattern discrimination is enhanced if location of target pattern is cued (Saarinen, 1993). A very interesting type of attentional enhancement was found by Yeshurun and Carrasco (1998). They found that the performance of human subjects decreased for a specific texture segregation task, if subjects

attended to it at the centre of gaze, but not at peripheral locations in the visual field. This texture segregation task was designed in a way that performance diminished, when spatial resolution was too high. Therefore, these results imply that visual attention enhances spatial resolution.

The neurophysiological correlate of attending to a spatial location is increased processing of the attended stimulus compared to the unattended one. In human studies, most researchers found attentional modulation of a peak in the event-related potential (ERP) starting around 80-100ms after stimulus onset (e.g. Hillyard, Vogel, & Luck, 1998). This so-called P1 peak is believed to reflect feed-back from higher visual areas to primary visual cortex (V1). However a recent study by Kelly, Gomez-Ramirez, and Foxe (2008) showed that even the first information arriving in V1 is modulated by spatial attention.

### **1.1.2 Object-based attention**

There are several experiments indicating that objects are another important target of visual attention. Briefly presenting two very small, overlapping objects, Duncan (1984) found that two judgments could be made in the same object, but not in different objects. Another psychophysical evidence for object-based attention comes from Blaser, Pylyshyn, and Holcombe (2000). They employed two Gabor patches which had different trajectories through a space spanned by the features color, orientation, and spatial frequency, but shared the same spatial location. Only based on their different trajectories in the feature space subjects were able to track one of the two objects and ignore the other one. On the electrophysiological side, Valdes-Sosa, Bobes, Rodriguez, & Pinilla (1998) showed that object-based attention modulates the P1 and N1 component of the ERP. They used two sets of dots in rigid rotation around the fixation spot, which created the percept of transparent surfaces. Subjects had to attend to one set of dots based on their color. The researchers found a strong suppression of the P1 and N1 component of a motion-elicited ERP for the unattended objects. Further physiological evidence for object-based attention stems from a functional Magnetic Resonance Imaging (fMRI) study by O'Craven, Downing, & Kanwisher (1999). They showed that attending to the motion of a face or a house enhanced the neural representation of the whole visual object, even though all visual objects occupied the same spatial location. These studies (and many others) provide evidence for objects as possible targets for visual attention. A recent study by Martinez and colleagues (2007) shed light on the interplay between space-based and object-based attention. These researchers found that directing spatial attention to one part of an object facilitates the processing of the entire object. This could be taken as indication of for a spread of attention throughout the boundaries of an object.

### **1.1.3 Feature-based attention**

In addition to spatial locations and objects, humans are also capable of attending to different features of an object like motion, color, or orientation. When subjects have to divide their attention between two spatial locations, subjects perform better on a dual-task when stimuli share a feature compared to when they do not (Saenz, Buracas, & Boynton, 2003). Using fMRI, the same authors found increased activity in brain areas that process a certain feature, only if this feature shared between the two spatial location (Saenz, Buracas, & Boynton, 2002). This indicates that visual attention can selectively enhance visual features.

## ***1.2 Overt and Covert Visual Attention***

Most researchers propose two different, but interacting mechanisms of visual attention: covert and overt visual attention. The former does not involve eye-movements and refers to a shift of attention without shift of gaze. In contrast, overt visual attention is related to eye-movements and directing the gaze at interesting locations. A theoretical framework unifying covert and overt visual attention is provided by the Pre-Motor theory of attention (Rizzolatti, Riggio, Dascola, and Umiltà, 1987). These authors propose the notion of a motor program that controls both types of attention. In this motor program the features like direction or distance of the movement are specified independently and in series.

### **1.2.1 Covert visual attention**

A first description of covert visual attention dates back to Herrmann von Helmholtz (1867). In his seminal book “Handbuch der physiologischen Optik”, he describes his subjective experience during a covert attention task using letters presented away from the point of fixation. He concluded that only by means of intentional control he could concentrate on a particular part of the visual field without eye-movements. Several more recent studies have confirmed and quantified these early observations (e.g. Posner, 1980). Another example for covert visual attention is the modulation of neuronal responses in primary visual cortex of macaque monkey in a mental curve-tracing task. Although the monkeys only mentally traced curves, the responses of neurons along the curves were modulated by task demands (Roelfsema et al, 2003). These results show that we can attend away from the centre of gaze.

### **1.2.2 Overt visual attention**

Overt visual attention is a more everyday mode of visual attention. We focus our gaze at objects or persons that seem interesting to us. However, overt visual attention is closely linked to covert visual attention. It has been shown that eye-movements and attention are

correlated in human subjects (Hoffman & Subramaniam, 1995; Maioli, Benaglio, Siri, Sosta, and Kappa, 2001). In animal experiments, cells in superior colliculus have been found, that are active during saccade preparation and covert shifts of attention (Kustov & Robinson, 1996; Ignashchenkova, Dicke, Haarmeier, and Their, 2004), indicating that there is also a common neuronal source. The analysis of eye-movements therefore provides an objective measure of attentional processes.

### ***1.3 Bottom-Up and Top-Down control of attention***

In the control of attention we distinguish two types of signals, those related to the stimulus (“bottom-up”) and those related to subject’s expectation and experience as well as to the task (“top-down”).

#### **1.3.1 Bottom-up**

The influence of bottom-up stimulus features on overt attention has been extensively studied in visual search tasks. In these tasks, participants have to search for a pre-defined target in an array of stimuli. The size (number of elements) of the array is varied by the experimenter. Another factor which can be changed is the similarity between target of the visual search and the other elements in the display, called distracters. If the target can be differentiated from the distracters based on a difference in one feature (e.g. red vs. green color), then reaction time of the subjects is independent of the set size. This indicates that features are able to “pop-out” in such a display. A likely explanation for this effect is the existence of an early parallel process of feature processing (Treisman & Gelade, 1980). However, there is a linear increase in reaction time with set size, if successful search depends on a conjunction of different features. This indicates that in conjunction search focal attention has to be serially applied to each relevant location. These notions have led to the formulation of the Feature-Integration Theory (Treisman & Gelade, 1980). However, there is criticism concerning the influence of the parallel process during serial search. Wolfe, Cave, and Franzel (1989) therefore postulated the Guided Search Model. In this model, the output of the parallel processing stage can be used by the serial stage during conjunction search. This model can also explain why searches for triple conjunctions (Color x Size x Form) are easier than for simple conjunctions. These models therefore predict a parallel feature processing stage, which is followed by a slower serial stage for more complex searches.

A neurobiologically inspired implementation of bottom-up driven visual attention is the so-called “saliency map” (Koch & Ullman, 1985). While implementations of this model underwent several modifications since its original formulation, the basic scheme remains

unchanged (Itti & Koch, 2001 for review). The stimulus is analyzed in various feature channels, such as luminance, color, or orientation. Local differences in each feature are computed, combined across different spatial scales, and normalized non-linearly. The resulting maps are summed, yielding the saliency map. Locations of high activity in this saliency map are assumed to be salient, i.e. likely to be attended. Recently, Peters, Iyer, Itti, and Koch (2005) performed a thorough comparison between the prediction of saliency map models and human fixations for a large variety of stimuli. These authors find that their models, especially versions that include additional interactions within the orientation channel, predict human fixation patterns reasonably well. This finding has been replicated recently with a simpler version of a saliency-map model (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.

In recent years, several studies examined the influence of individual stimulus features on human overt attention in natural scenes. For the arguably best investigated feature, luminance-contrast, Mannan, Ruddock, and Wooding (1996) find that luminance-contrast is elevated at fixation locations, if the analysis is corrected for general biases in stimuli and fixation patterns. Using a different definition of luminance-contrast, Reinagel and Zador (1999) also report this correlative effect between luminance-contrast and human overt attention. This result was later confirmed in several other studies (Krieger, Rentschler, Hauske, Schill, & Zetzsche, 2000; Parkhurst & Niebur, 2003), although under some presentation conditions a dependence on spatial frequency may obstruct the effect for unfiltered stimuli (Mannan, Ruddock, & Wooding, 1997; Tatler, Baddeley, & Gilchrist, 2005). Recently, a study of our laboratory also replicates the correlation of luminance-contrast to human fixation, but demonstrates that this effect of luminance-contrast is not causal (Einhäuser & König, 2003). Hence, first-order luminance-contrast – despite being elevated at fixations - is not causally related to the bottom-up control of human overt attention. We therefore have to pay special attention to correlative effects when analyzing the salience of features.

### **1.3.2 Top-down**

The interest in top-down features guiding human fixations dates at least back to Buswell's (1935) study. Comparing students of art to average observers, he finds fixation durations to be slightly shorter for the "expert" group when looking at pictures; during reading (but not for pictures) a similar difference held true between adults and children. Both results highlight the

role of subjective experience and special training to human attention. In addition, Buswell provides some qualitative account of the effect of the task, which three decades later Yarbus' (1967) seminal study addresses in detail. Yarbus, in particular, demonstrates the dramatic influence of the task on overt attention in scenes that contain objects of high behavioral relevance to human observers, such as people or faces.

In visual search tasks, top-down control has been examined by making subjects adopt certain strategies, which proved to be resistant to bottom-up influences (Folk, Remington, and Johnston, 1992; Bacon and Egeth, 1994). Based on such results, there seems to be general consensus on the relative importance of experience and task to human attention.

Nevertheless, bottom-up features are also assumed to play an important role for the allocation of spatial attention, especially in the absence of an explicit task. Consequently, the investigation of bottom-up models and their interaction with top-down signals provides fruitful insight in the mechanisms underlying human attention.

### **1.4 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 535nm), and long (L, 565nm). 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). The outputs of the retinal ganglion cells are processed by two color opponent mechanisms in the parvocellular and koniocellular 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. Further color processing takes place in chromoxidase-rich blobs of the primary visual cortex (V1), area V4 (Zeki, 1983), and area IT (Komatsu, Ideura, Kaji, & Yamane, 1992). A recent study by Mazer and Gallant (2003) of macaque monkeys viewing grayscale natural scenes indicates that ventral area V4 and IT are involved in the computation of saliency. In area V4 parallel and serial search mechanism are

present during visual search for color and/or shape defined targets (Bichot, Rossi & Desimone, 2005). These neurophysiological studies provide evidence for attentional mechanisms in brain areas involved in color processing.

On the psychophysical side, there are several studies on the influence of color on visual attention. In visual search experiments the influence of color on attention has been studied in great detail. Color singletons “pop-out” in large search displays, i.e. the reaction time does not depend on the size of the search display (Treisman & Gelade, 1980; Nagy & Sanchez, 1990, D’Zmura, 1991). Color singletons capture attention even irrespective of the observers’ attentional set (Theeuwes, 1994). However, if the difference in hue between distracters and target is small (e.g. green distracters and yellowish-green target), the reaction time will depend on the search display size (Nagy & Sanchez, 1990). This indicates a boundary for pop-out effects of color in visual search.

Another important line of research is concerned with local operators in the color domain that predict fixation locations in natural color images. It was shown that chromaticity is a predictive feature (Tatler, Baddeley, & Gilchrist, 2005), but its salience differs between image categories (Parkhurst & Niebur, 2002). However, we can describe color information by several physiologically plausible features, like saturation or color-contrast in the R-G and B-Y opponent color channels. These features influence the firing of neurons in the retino-geniculate channels as well as parvocellular and koniocellular layers of LGN (Hendry & Reid, 2000; Gegenfurtner & Kiper, 2003). Therefore, current literature still lacks a thorough examination of the salience of neurobiologically plausible color features in natural scenes.

### ***1.5 Aim of this thesis***

There are only few studies examining the influence of color on overt visual attention in natural scenes. Therefore, the aim of this thesis is to evaluate the influence of neurophysiologically plausible color features on overt visual attention. Since the analysis of features at fixated image locations can only reveal a correlation of features with overt attention, I use several color-modified versions of the same image. These modifications allow me to examine the salience of color features beyond correlational observations.

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

Hans-Peter Frey, Christian Honey, & Peter König (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.

### **2.0 Abstract**

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 (Itti, Koch, and 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.

### **2.1 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 (Kustov & Robinson, 1996; Ignashchenkova, Dicke, Haarmeier, & Thier, 2004), 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 towards 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 535nm), and long (L, 565nm). 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 (see Itti & Koch, 2001 for a review). 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 & 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 2<sup>nd</sup> 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 & Niebur, 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, Krauskopf, & Lennie, 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.

## **2.2 Methods**

### **2.2.1 Subjects**

15 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 (Kanehara 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.

### **2.2.2 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, König, & Einhäuser, 2007). One such color space is DKL space (Derrington, Krauskopf, & Lennie, 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.

### **2.2.3 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 (Açık 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 dataset (Troscianko, Párraga, Leonards, Baddeley, Troscianko, & Tolhurst, 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 2.1 (panel A-G).

The images were down-sampled to a resolution of 1024×768 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 2.1 (panel H). For stimulus presentation, we used a 21'' 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 28×21 degrees of visual angle for our stimuli.

#### **2.2.4 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 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  $8000^{\circ}/s^2$ , the velocity was higher than  $30^{\circ}/s$ , 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.



Figure 2.1 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

### 2.2.5 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.

### 2.2.6 Definition of features

We analyzed the influence of two luminance features and three color features on subjects' fixation behavior: luminance contrast, texture contrast (2<sup>nd</sup> 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 doesn't 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 and König, 2003, Frey, König, and Einhäuser, 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.

### 2.2.7 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 Tatler, Baddeley, & Gilchrist, 2005; Frey, König, & Einhäuser, 2007).

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 2.2, 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 2.2, 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  $p < 0.01$  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  $1 - \text{specificity}$  (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.

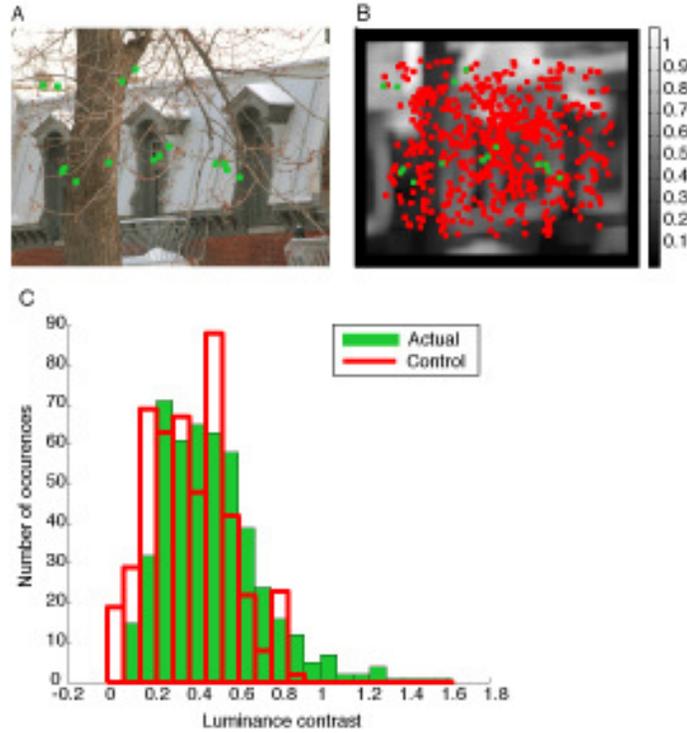


Figure 2.2 Feature analysis.

(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  $p < 0.01$ . The ROC AUC value is 0.62. For presentation, the distributions are binned using 20 bins.

### 2.2.8 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

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

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 (Peters et al, 2005; Le Meur, Le Callet, and Barba, 2007) 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 2.3).

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.

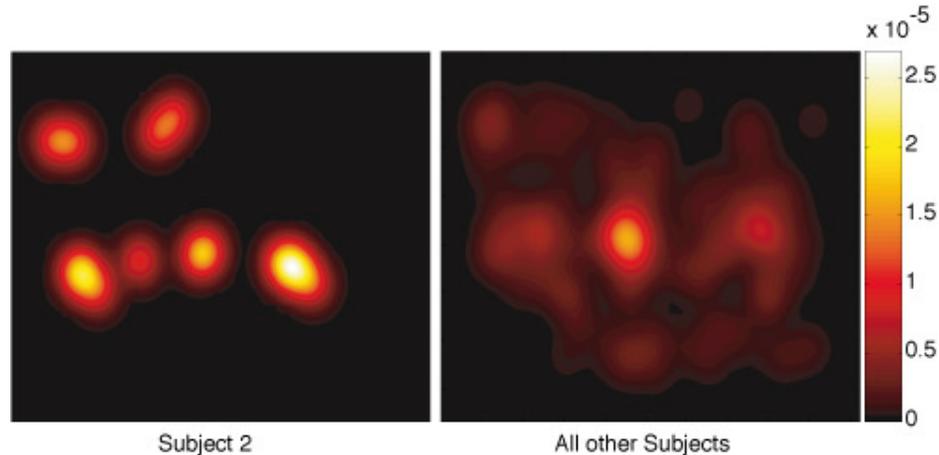


Figure 2.3 Calculation of congruency between observers.

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.

### 2.2.9 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, Koch, & Niebur (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.

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.

### **2.2.10 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" 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.

## **2.3 Results**

### **2.3.1 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 towards 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 2.4, panel A). For these four categories we can assume an influence of color on overt attention.

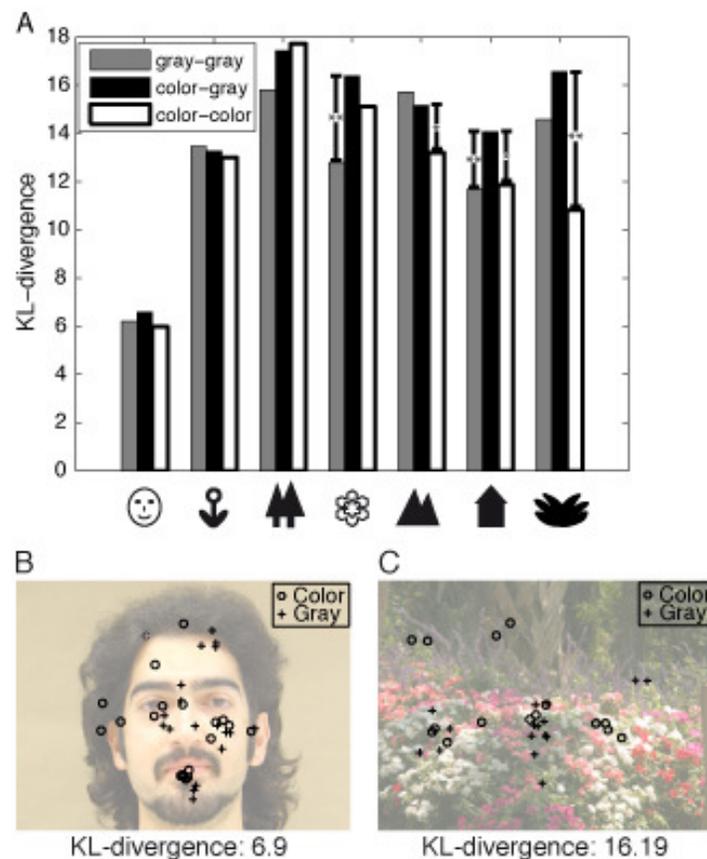


Figure 2.4 Similarity of fixation locations in colored and grayscale conditions.

A: Mean KL-divergence 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.

In *Face*, *Flower*, and *Landscape* 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 2.4, 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. 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 2.5, panel A). In all other categories the congruency is significantly lower, with the highest KL-divergence in *Forest* (17.9; Figure 2.5 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 2.5, panel B), meaning that the color information in *Rainforest* images leads subjects to fixate 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 2.5, 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*.

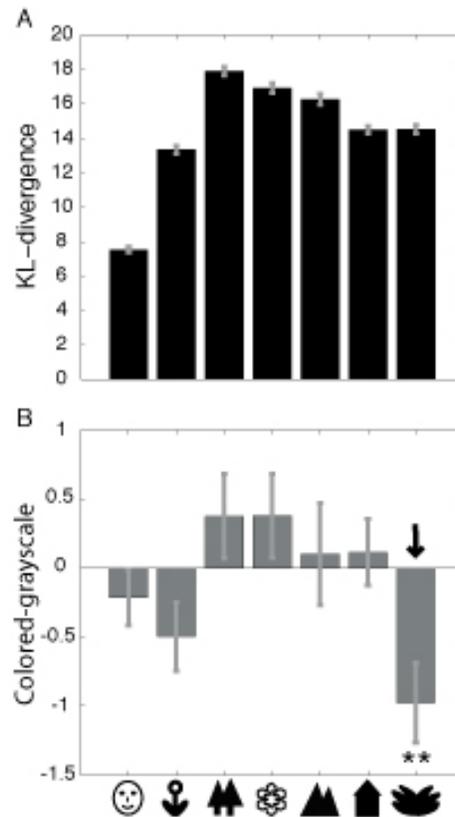


Figure 2.5 Inter-observer congruency.

A: Mean KL-divergence (with standard error of mean [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.

## 2.3.2 Stimulus Features

### 2.3.2.1 Luminance features

Fixated regions differ from non-fixated ones with respect to several image features. In grayscale images, luminance contrast or 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.

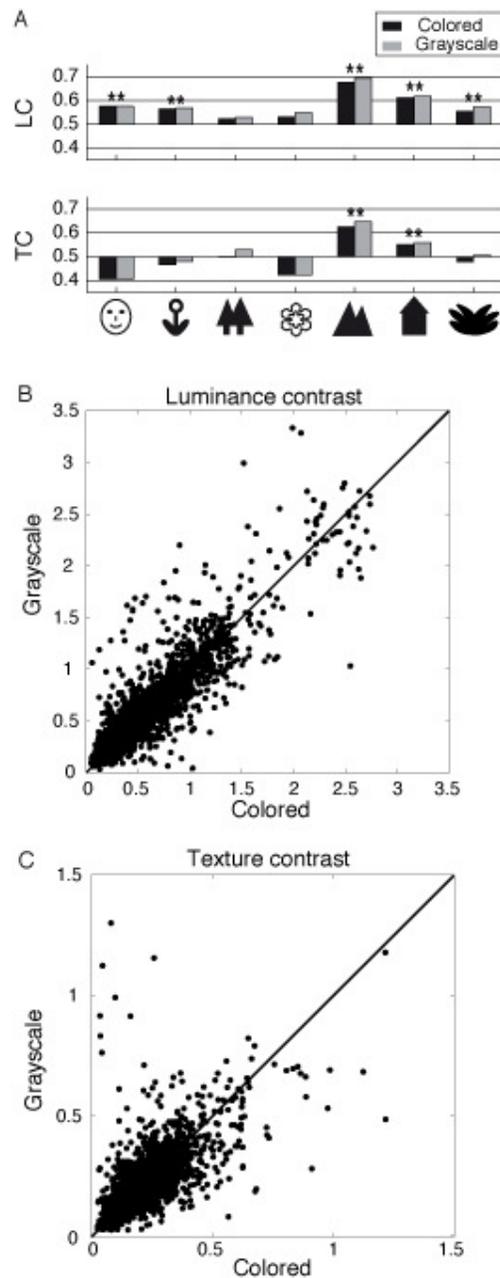


Figure 2.6 Luminance features.

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.

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 2.6, 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 2.6, 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 2.6 B and C, 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 salience of luminance features.

### 2.3.2.2 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 salience 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 2.7, 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 2.7, 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.

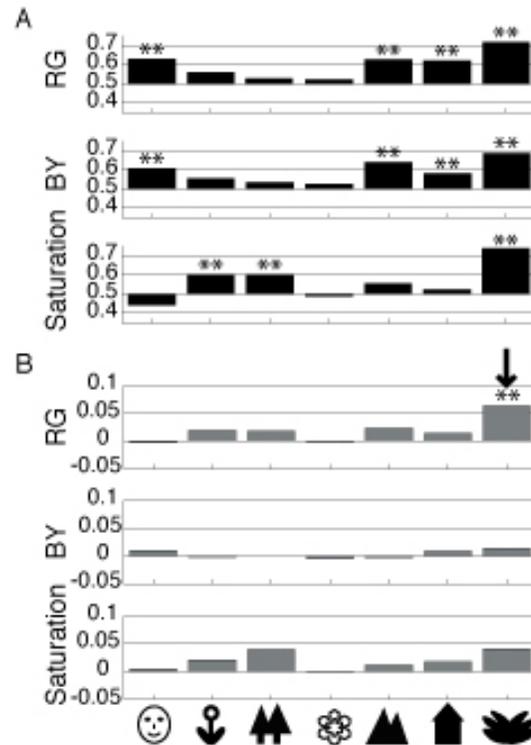


Figure 2.7 Color features.

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.

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 salient* if it guides our attention not only by virtue of such a coincidental correlation. To assess the original salience 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 2.7, 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 2.9), 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.

### 2.3.3 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.

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

Table 2.1 Influence of color on the saliency map model.

Mean KL-divergence between saliency maps for colored and grayscale versions of the same image.

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 2.8, 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 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?

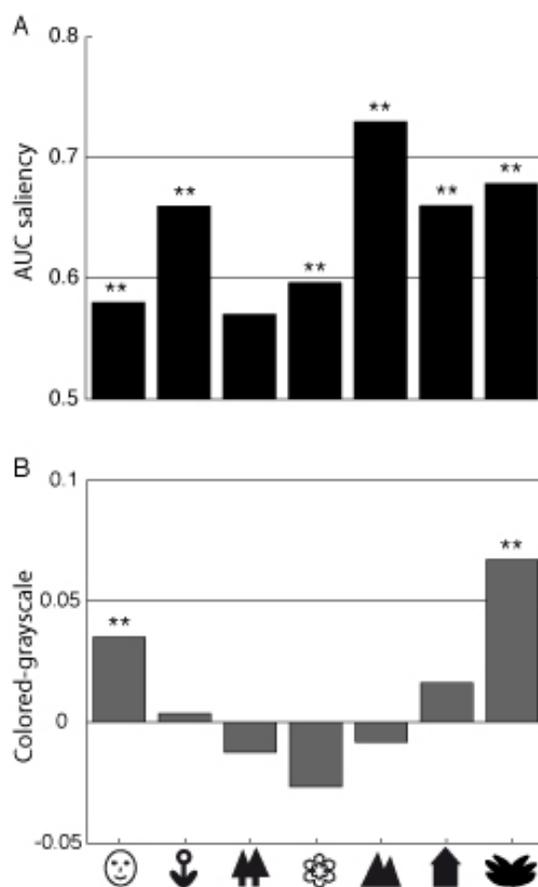


Figure 2.8 Saliency map model.

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.

AUC values are significantly higher in colored compared to grayscale images in *Face* and *Rainforest* ( $p < 0.01$ , KS-test with Bonferroni correction; Figure 2.8, 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, and 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).

### 2.3.4 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.

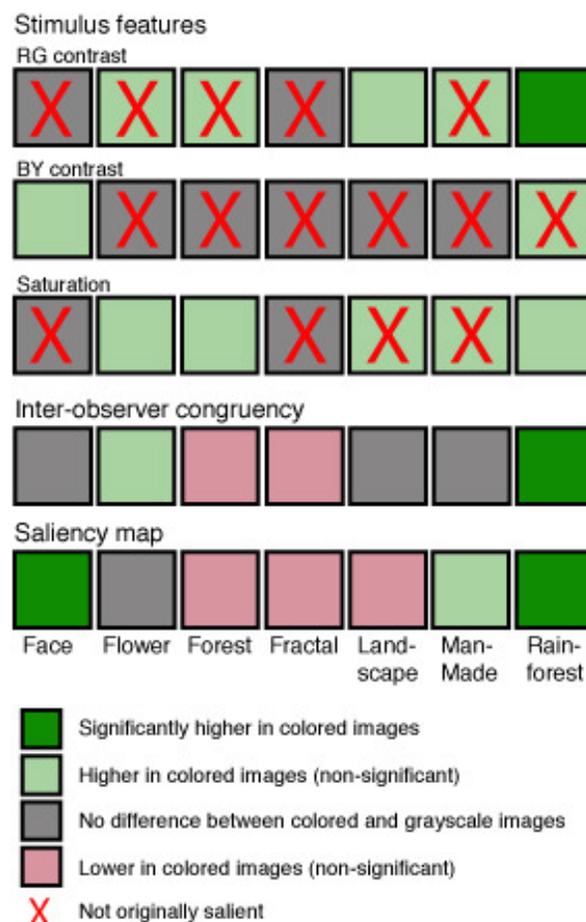


Figure 2.9 Influence of color on humans and saliency map.

Dark green indicates significantly higher values in colored compared to grayscale images. Light green/ red represent 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.

## **2.4 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.

### **2.4.1 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 locations in colored and grayscale version of *Face* stimuli are similar. Color information doesn't make subjects look at different locations. This parallels findings in face recognition tasks. Kemp, Pike, White, & 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.

### 2.4.2 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, & Koch (2008). 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.

### 2.4.3 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 towards a strong influence of RG contrast on overt attention in this category.

### 2.4.4 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. Tatler, Baddeley, & Gilchrist, 2005; Frey, König, & Einhäuser, 2007) 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.

### 2.4.5 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, and Desimone, 2005; for a review see: Desimone and 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, Law, & Niebur, 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 (Hendry & Reid, 2000; Gegenfurtner & Kiper, 2003).

Using these detailed color features in our analysis, we find different results than earlier studies. Contrary to Parkhurst, Law, & Niebur (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, Baddeley & Gilchrist

(2005) for images that could be attributed to the *Man-Made* category. Here we can not 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 can't 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 and 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.

## **2.5 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.
- Buswell, G. T. (1935). How people look at pictures; a study of the psychology of perception in art. Chicago, Ill: University of Chicago Press.
- Carmi, R., & Itti, L. (2006). Visual causes versus correlates of attentional selection in dynamic scenes. *Vision Res*, 46, 4333-4345.
- Cerf, M., Harel, J., Einhäuser, W., & Koch, C. (2008). Predicting human gaze using low-level saliency combined with face detection. In A. McCallum (Ed.), *Advances in Neural*

- Information Processing Systems (NIPS) (in press).
- Changizi, M. A., Zhang, Q. & Shimojo, S. (2006). Bare skin, blood, and primate colour vision. *Biol Lett*, 2(2), 217-221.
- Dayan, P., & Abbott, L. F. (2001). *Theoretical neuroscience: computational and mathematical modeling of neural systems*. Cambridge, Mass.: Massachusetts Institute of Technology Press.
- Derrington, A. M., Krauskopf, J., & Lennie, P. (1984). Chromatic mechanisms in lateral geniculate nucleus of macaque. *J Physiol*, 357, 241-265.
- Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annu Rev Neurosci*, 18, 193-222.
- Dominy, N. J., & Lucas, P. W. (2001). Ecological importance of trichromatic vision to primates. *Nature*, 410, 363-366.
- Einhäuser, W., & König, P. (2003). Does luminance-contrast contribute to a saliency map for overt visual attention? *Eur J Neurosci*, 17, 1089-1097.
- Faraggi, D., & Reiser, B. (2002). Estimation of the area under the ROC curve. *Stat Med*, 21, 3093-3106.
- Felsen, G., & Dan, Y. (2005). A natural approach to studying vision. *Nat Neurosci*, 8, 1643-1646.
- Frey, H. P., König, P., & Einhäuser, W. (2007). The role of first- and second-order stimulus features for human overt attention. *Percept Psychophys*, 69, 153-161.
- Gegenfurtner, K. R., & Kiper, D. C. (2003). Color vision. *Annu Rev Neurosci*, 26, 181-206.
- Gottlieb, J. P., Kusunoki, M., & Goldberg, M. E. (1998). The representation of visual salience in monkey parietal cortex. *Nature*, 391, 481-484.
- Helmholtz, H. von (1867). *Handbuch der physiologischen Optik*. Hamburg: Verlag Leopold Voss.
- Hendry, S. H. C., & Reid R. C. (2000). The koniocellular pathway in primate vision. *Annu Rev Neurosci*, 23, 127-153.
- Hoffman, J. E., & Subramaniam, B. (1995). The role of visual attention in saccadic eye movements. *Percept Psychophys*, 57, 787-795.
- Horwitz, G. D., & Newsome, W. T. (1999). Separate signals for target selection and movement specification in the superior colliculus. *Science*, 284, 1158-1161.
- Ignashchenkova, A., Dicke, P. W., Haarmeier, T., & Thier, P. (2004). Neuron-specific contribution of the superior colliculus to overt and covert shifts of attention. *Nat Neurosci*, 7, 56-64.

- Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nat Rev Neurosci*, 2, 194-203.
- Itti, L., Koch, C., & Niebur, E. (1998). A Model of Saliency-Based Visual Attention for Rapid Scene Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11), 1254-1259.
- 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.
- 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. *Hum Neurobiol*, 4, 219-227.
- Komatsu, H., Ideura, Y., Kaji, S., & Yamane, S. (1992). Color selectivity of neurons in the inferior temporal cortex of the awake macaque monkey. *J Neurosci*, 12, 408-424.
- Kullback, S., & Leibler, R. A. (1951). On Information and Sufficiency. *The Annals of Mathematical Statistics*, 22(1), 79-86.
- Kustov, A. A., & Robinson, D. L. (1996). Shared neural control of attentional shifts and eye movements. *Nature*, 384, 74-77.
- Le Meur, O., Le Callet, P., & Barba, D. (2007). Predicting visual fixations on video based on low-level visual features. *Vision Res*, 47, 2483-2498.
- Li, A., & Lennie, P. (2001). Importance of color in the segmentation of variegated surfaces. *J Opt Soc Am A Opt Image Sci Vis*, 18, 1240-1251.
- 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. *Eur J Neurosci*, 13, 364-372.
- 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. *Spat Vis*, 10, 165-188.
- 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.
- Nathans, J. (1999). The evolution and physiology of human color vision: insights from molecular genetic studies of visual pigments. *Neuron*, 24, 299-312.

- Olmos, A., & Kingdom, F. A. A. (2004). McGill Calibrated Colour Image Database (<http://tabby.vision.mcgill.ca>).
- Parkhurst, D., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision Res*, 42, 107-123.
- Parkhurst, D. J., & Niebur, E. (2004). Texture contrast attracts overt visual attention in natural scenes. *Eur J Neurosci*, 19, 783-789.
- Peters, R. J., Iyer, A., Itti, L., & Koch, C. (2005). Components of bottom-up gaze allocation in natural images. *Vision Res*, 45, 2397-2416.
- Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annu Rev Neurosci*, 13, 25-42.
- Reinagel, P., & Zador, A. M. (1999). Natural scene statistics at the centre of gaze. *Network*, 10, 341-350.
- Sumner, P., & Mollon, J. D. (2000). Catarrhine photopigments are optimized for detecting targets against a foliage background. *J Exp Biol*, 203, 1963-1986.
- Tatler, B. W., Baddeley, R. J., & Gilchrist, I. D. (2005). Visual correlates of fixation selection: effects of scale and time. *Vision Res*, 45, 643-659.
- Thompson, K. G., Bichot, N. P., & Schall, J. D. (1997). Dissociation of visual discrimination from saccade programming in macaque frontal eye field. *J Neurophysiol*, 77, 1046-1050.
- 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 (ECP Abstract Supplement).
- Zeki, S. (1983). The distribution of wavelength and orientation selective cells in different areas of monkey visual cortex. *Proc R Soc Lond B Biol Sci*, 217, 449-470.

### **3. A hue invariant causal influence of color on eye movements and a high level compensatory mechanism while studying images of rainforest.**

Hans-Peter Frey, Kerstin Wirz, Verena Willenbockel, Torsten Betz, Cornell Schreiber, Tomasz Troscianko, & Peter König (submitted).

#### ***3.0 Abstract***

Recent research indicates a correlation of color features on visual attention under natural conditions. Building on these results we here examine the causal influence of color features on overt attention. To this end we recorded eye-movements of color-normal and deuteranope human subjects freely viewing Rainforest images. These images were either presented in natural color or parametrically modified. Consistent with previous results we observe in color-normal subjects a high correlation of selected fixation points with red-green contrast and a low correlation with blue-yellow contrast. Eliminating red-green color information dramatically alters fixation behavior, indicating for a causal effect of color contrast. However, if rotating the hue in these images, no correlation is observed of red-green color information with selected fixation points. A detailed analysis uncovers that this results of a mechanism analyzing color contrast invariant of a specific axis in color space interacting with the chromatic properties of natural images. In deuteranope participants we found surprisingly a correlation of red/green contrast with overt attention comparable in strength to normal subjects. However, the time dependence indicates that this is not a fast direct effect, but a high-level compensatory mechanism. Together these results argue for a causal influence of color contrast on overt attention in natural stimuli, processed in a way that it is not specific to the red-green axis in hue space, and a slower high-level compensatory mechanism in the case of color blindness.

#### ***3.1 Introduction***

The ability to perceive colors is a decisive component of human vision. Color helps us to segment visual scenes, recognize, and memorize objects. For more than one century, researchers have examined the components and pathways of the color system. In this paper we utilize the knowledge about color processing in order to assess its influence on visual attention.

### 3.1.1 Visual Attention

Typically researchers distinguish between covert and overt attention. The former does not involve eye-movements and refers to a shift of attention without shift of gaze and a first description of this phenomenon dates back to Helmholtz (1867). The latter is related to eye-movements and directing the gaze at interesting locations. However, it has been shown that eye-movements and attention are closely correlated in human subjects (Hoffman & Subramaniam, 1995; Maioli, Benaglio, Siri, Sosta, and Kappa, 2001). In animal experiments, cells in superior colliculus have been found, that are active during saccade preparation and covert shifts of attention (Kustov & Robinson, 1996; Ignashchenkova, Dicke, Haarmerier, and Their, 2004), indicating that there is also a common neuronal source. The analysis of eye-movements therefore provides an objective measure of attentional processes.

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). In visual search tasks, top-down control has been examined by making subjects adopt certain strategies, which proved to be resistant to bottom-up influences (Folk, Remington, and Johnston, 1992; Bacon and Egeth, 1994). However, several other researchers have shown that highly salient singleton features are able to override this top-down control of attention (Yantis and Yonides, 1990; Theeuwes, 1992). This dichotomy also exists about the relative timing of these two forms of attentional control: some researchers found bottom-up influences to operate an earlier time-point than top-down influences (e.g. Müller and Rabbit, 1989; Cheal and Lyon, 1991; van Zoest, Donk, and Theeuwes, 2004), while other studies could not find evidence (e.g. Tatler, Baddeley, and Gilchrist, 2005). Therefore the question when and how bottom-up and top-down processes influence overt attention remains vividly disputed.

In eye-tracking studies examining natural scenes several possible bottom-up features were analyzed. 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), intrinsic dimensionality i2D features, like curved edges or corners (Krieger, Rentschler, Hauske, Schill, & Zetsche, 2000; Saal, Nortmann, Krüger, and König, 2006) or texture contrast (2<sup>nd</sup> order luminance contrast; Parkhurst & Niebur, 2004). 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; Aık et al, submitted). Therefore we have to pay special attention to correlative effects when analyzing the salience of features.

### 3.1.2 Color processing

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 535nm), and long (L, 565nm). The color of an object can only be computed unambiguously if the magnitudes of the outputs of all three cone types are compared. In macaque monkeys, 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 parvo- and koniocellular 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 retinal ganglion cells in the LGN. These findings 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 M- and L-cones (Nathans, 1999). For ease of reading we will refer to this color-channel as the R-G channel. Another subsystem compares the output of the S-cones with the combination of L- and M-cones (Derrington, Krauskopf, and Lennie, 1984). We will refer to this color channel as the B-Y channel. 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; Regan et al, 2001) or young leaves (Sumner & Mollon, 2000; Dominy & Lucas, 2001). 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. Previously we analyzed the image statistics of color-calibrated images out of this rainforest (Frey, Wirz, Willenbockel, Troscianko & K nig, 2006). We found that R-G color-contrast (R-G contrast) in an area of the size of the fovea is significantly higher at fruits compared to all other fixation locations. No such effect was found for B-Y color-contrast (B-Y contrast). In addition, examining only the excitations in the opponent color-channels did not reveal any difference, either. This indicates that fruit objects in Rainforest images produce very high R-G contrast.

### 3.1.3 Color and attention

Further color processing takes place in chromoxidase-rich blobs of the primary visual cortex (V1), area V4 (Zeki, 1983), and area IT (Komatsu, Ideura, Kaji, & Yamane, 1992). A recent study by Mazer and Gallant (2003) of macaque monkeys viewing grayscale natural scenes indicates that ventral area V4 and IT are involved in the computation of saliency. In area V4 parallel and serial search mechanism are present during visual search for color and/or shape defined targets (Bichot, Rossi & Desimone, 2005). These neurophysiological studies provide evidence for attentional mechanisms in brain areas involved in color processing.

On the psychophysical side, there are several studies on the influence of color on visual attention.

In visual search experiments the influence of color on attention has been studied in great detail. Color singletons “pop-out” in large search displays, i.e. the reaction time does not depend on the size of the search display (Treisman & Gelade, 1980; Nagy & Sanchez, 1990, D’Zmura, 1991). Color singletons capture attention even irrespective of the observers’ attentional set (Theeuwes, 1994). However, if the difference in hue between distracters and target is small (e.g. green distracters and yellowish-green target), the reaction time will depend on the search display size (Nagy & Sanchez, 1990). This indicates a boundary for pop-out effects of color in visual search.

Analyzing the time-course attentional selection, it has been shown that stimulus driven and goal driven control occur at different time points (e.g. Muller & Rabbitt, 1989; van Zoest, Donk & Theeuwes, 2004).

Another important line of research is concerned with local operators in the color domain that predict fixation locations in natural color images. It was shown that chromaticity is a predictive feature (Tatler, Baddeley, & Gilchrist, 2005), but its salience differs between image categories (Parkhurst & Niebur, 2002; Frey, Honey, & König, 2008). However, we can describe color information by several physiologically plausible features, like saturation or color-contrast in the R-G and B-Y opponent color channels. These features influence the firing of neurons in the retino-geniculate channels as well as parvocellular and koniocellular layers of LGN (Hendry & Reid, 2000; Gegenfurtner & Kiper, 2003). Using these color features, we have shown that different color subsystems influence overt visual attention selectively in certain environments. While saturation is the local color operator that predicts best where we fixate in images of flowers and forests, color-contrasts predict best in images of European or Northern American landscapes as well as rainforest (Frey, Honey, & König, 2008). The category, in which we found the strongest influence of color-contrasts on overt

attention, is Rainforest. In order to examine a possible causal influence of color features on overt attention, we therefore chose Rainforest images.

### **3.1.4 This study**

In this study, we examine whether there is a causal relationship between color features and overt attention in Rainforest images. To this end, we employ 2 different experimental stimulus manipulations: reduction and manipulation of color information. These manipulations allow us to examine the influence of color features on overt visual attention beyond correlational observations. The manipulated stimuli are presented to color-normal subjects and deuteranope subjects. The altered cone sensitivities in deuteranope subjects mostly affect the R-G color opponent channel, i.e. they have difficulties discriminating yellowish red from green hues. Examining the saliency of color features and fixation behavior in these two populations of subjects, allows us to distinguish between causal stimulus-driven and higher-level influences.

## **3.2 Experiment 1**

### **3.2.1 Introduction**

In Experiment 1, we reduce the chromaticity along each of the two cardinal color axes. This yields images either completely devoid of either R-G or B-Y color information. In each condition, our modification does not affect the remaining two axes of the DKL color space. Experiment 1 therefore is the most straightforward approach to examine a causal influence of color channels on overt visual attention.

### **3.2.2 Methods**

#### **3.2.2.1 Color representation**

In this study, the choice of the color representation is decisive. Color images can be represented using either neurophysiologically or psychophysically defined color-spaces. We argue that neurophysiologically defined spaces are more appropriate for this purpose (Frey, König, & Einhäuser, 2007). Therefore we used the DKL color space (Derrington, Krauskopf, & Lennie, 1984) in all experiments.

This color space 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 (R-G) axis. 2. ‘Tritanopic

confusion' is defined by  $(L+M)-S$ . We will refer to this axis as the blue-yellow (B-Y) 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  $R-G > 0$ ,  $B-Y = 0$ ). 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.

### 3.2.2.2 Stimuli

Photographs of Uganda rainforest and color-modified versions thereof served as experimental stimuli. These images were all from the Kibale Forest image dataset (Troscianko, Párraga, Leonards, Baddeley, Troscianko, Tolhurst, & Troscianko 2003). They were obtained with a carefully color-calibrated digital camera (Nikon E950). The photographs depicted mainly leaves, fruit, bushes, and trees of Kibale Forest in variable distance. No man-made objects were present in the scenes. All color- modifications were carried out in DKL color space.

For stimulus presentation, we used a 21" 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 60 cm from the monitor surface, which yielded approximately  $37 \times 28$  degrees of visual angle for our stimuli.

In Experiment 1 we employed 40 images. Each of these images was presented in three different conditions. First, we used the unmodified DKL stimuli in which information of the full color spectrum is included. We will refer to these natural color stimuli as 'NAT' condition. Second, we set the saturation of the B-Y axis to zero, thus creating stimuli in which there is only luminance and red-green color information present. This condition will be referred to as 'noBY' condition. Third, we eliminated the influence of the R-G axis by setting the saturation of this axis to zero. We will refer to these stimuli as 'noRG'. One example image in all three conditions is depicted in Figure 3.1.

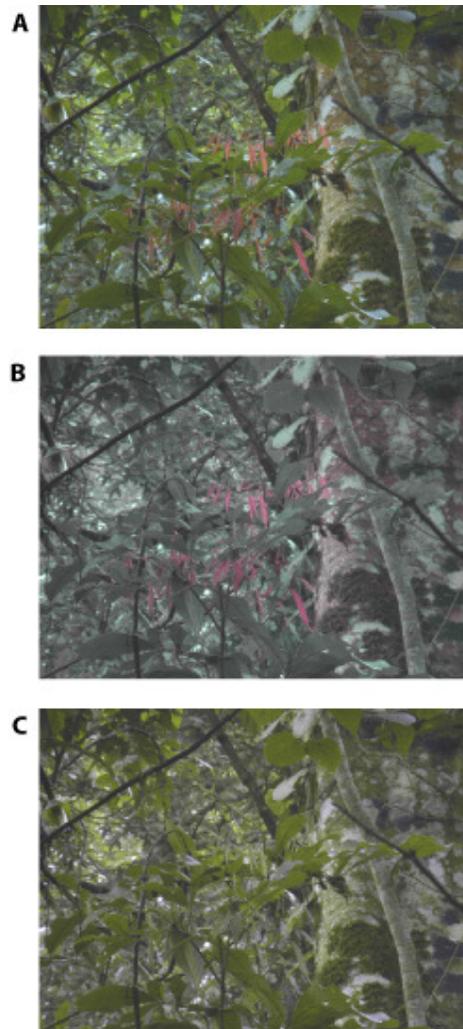


Figure 3.1 Experimental conditions in Experiment 1.

The same image is shown in conditions NAT (A), NoBY (B), and NoRG (C).

### 3.2.2.3 Subjects

15 undergraduate students from the University of Osnabrück participated in Experiment 1. 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 (Kanehara 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.

### 3.2.2.4 Experimental Design

In all experiments 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.

Experiment 1 was conducted in a sessions with 2 blocks of 60 images. The order of presentation was randomized for each subject, with the constraint that no image was presented in a row.

### **3.2.2.5 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 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  $8000^\circ/s^2$ , the velocity was higher than  $30^\circ/s$ , 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.

### **3.2.2.6 Definition of features**

We analyzed the influence of three color features on the subjects’ overt attention. These features are saturation, R-G, and B-Y color contrast. All these three features are neurophysiologically relevant.

The two color contrasts were defined solely as the standard deviation of the chromatic content of an image patch. We did not normalize by a mean color value of the image, as in the luminance case: A symmetrical distribution of color values along any color axis of DKL space would lead to a mean color value of 0. Narrow, symmetrical distributions with a mean close to zero would yield very high color contrast values we just don’t see in the color content of the image.

In DKL color space the saturation of a pixel is represented by the absolute value of the pixel’s

chromatic content. The saturation in an image patch therefore was defined as the mean saturation of all pixels in that patch.

Feature values are computed in a  $2.3^\circ$  square patch around a given pixel. Alternative patch sizes, ranging from  $1.2^\circ$  to  $4.5^\circ$ , were also used for all features but did not lead to any qualitative difference in the results.

### 3.2.2.7 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 a “central bias” (see Tatler et al., 2005; Frey et al, 2007). In each trial, the feature values at fixation locations are calculated as described above.

For each feature we define the “actual” value of that feature as the median of the feature values over all fixations (green crosses in Figure 3.2, panel A) in a trial ( $mA$ ). Each actual value was compared with a corresponding baseline that took into account potential biases in the subjects’ eye positions. Therefore, we defined as control fixations all fixations of the same subject on all other images in the same category and condition (red crosses in Figure 3.2, panel A). The median feature values at these control locations were the “control” values ( $mC$ ). The actual value would be different from the control value if and only if the feature had an effect on overt attention. These values were not normally distributed (Figure 3.2, panel B). Therefore we tested the significance of this difference using a non-parametric statistical test, the two-sided Kolmogorov-Smirnov test. Since we did several comparisons for each feature we Bonferroni corrected the p-values of the test. Features for which actual and control distributions differed with  $p < 0.01$  were termed salient.

In order to compare the differences between actual and control values among different

conditions we calculate a normalized effect size, defined as:  $eff = \frac{mA - mC}{mC}$ .

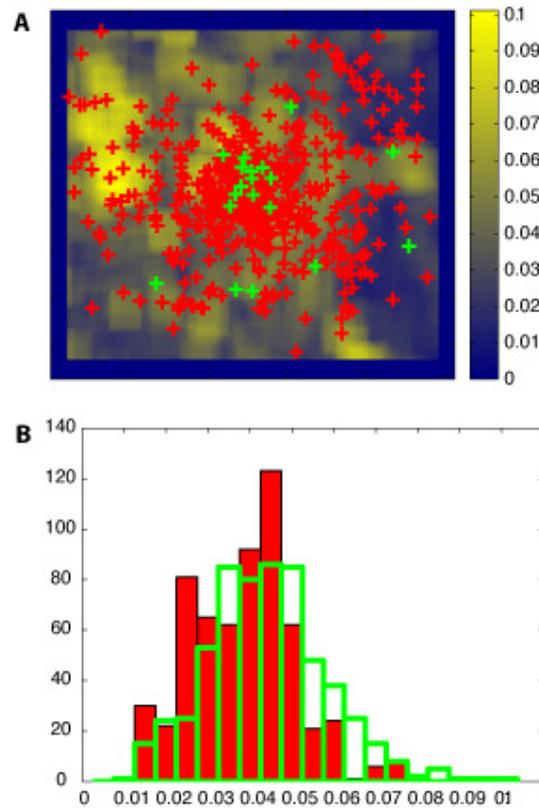


Figure 3.2 Feature analysis.

(A) Measured fixation locations of one subject (green) and corresponding control locations (red, see text for details) plotted on the B-Y color contrast map of the image in Figure 3.1A. For each image and subject, the actual value is defined as the median feature value over all fixations. Control values are defined in an analogous manner. (B) The distribution of actual (opaque bars with green edge) and control (red bars) B-Y color contrast for all subjects and images in the NAT condition. The KS-test indicates that these two distributions are significantly different with  $p < 0.01$ . The effect size is 0.19. For presentation only, the distributions are binned using 20 bins.

### 3.2.2.8 Congruency of fixation locations between conditions and observers

The design of our study allows us to determine an influence of color channels on overt attention by looking at the distribution of fixations. If fixations of a subject differ between the three conditions of the same image, then a certain aspect of color information influences overt attention. Examining the congruency between different observers allows us to further quantify this effect. 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

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

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 true 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. 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 3.3).

To determine the congruency between fixation locations on different versions of the same image seen by the same subject, we employ the same 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.

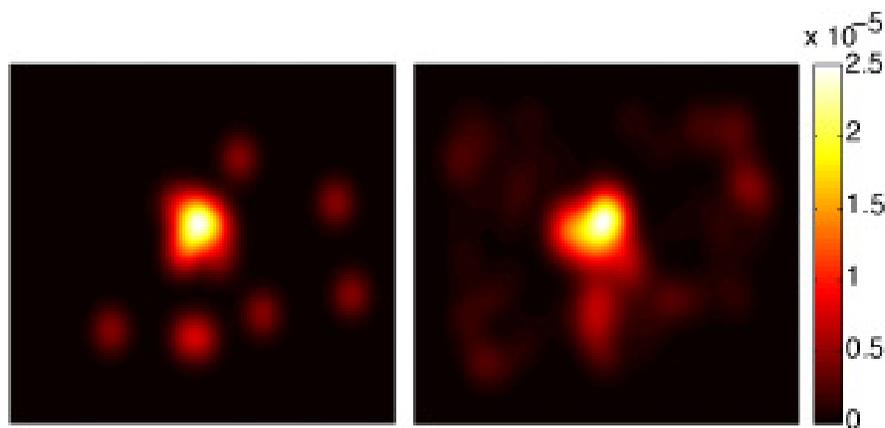


Figure 3.3 Calculation of congruency between observers.

We create a fixation probability map for each subject (left) as well as for all other subjects except the actual one (right). These two probability distributions are then compared using the Kullback-Leibler divergence. In this example, the KL-divergence is 17.19 bits.

### 3.2.3 Results

#### 3.2.3.1 Analysis of image features

This experiment is designed to assess the influence of the two cardinal color channels on overt visual attention. We expect that our experimental manipulation strongly influences overt visual attention. However, before we look at the actual eye-movements, we first analyze the saliency of color features at fixated locations.

The effect sizes for R-G contrast in images, which contain R-G color information, are very

high. For naturally colored we get a mean effect size of 0.65 and for NoBY 0.69 (Figure 3.4, upper panel). These values indicate that R-G contrast is 65% and 69% higher at fixated than at control fixations, respectively. The effect sizes for B-Y contrast are significantly smaller than those for R-G contrast in Natural images ( $p < 0.01$ ; two sided KS-test). The mean effect size is 0.16 in naturally colored and R-G reduced images. These results confirm our previous finding that R-G contrast is highly salient in Rainforest images, whereas B-Y contrast is rather not.

Virtual color contrasts, e.g. calculating R-G contrast on an image which subjects saw devoid of R-G color information, allow us to further elucidate the salience of color features. If a feature is only has a high effect size by a correlation with other features of the image then its virtual effect size should not be changed. This is exactly what we find for B-Y contrast (Figure 3.4, middle panel). However, we detect a significant drop of 0.2 in R-G contrast effect size in images devoid of R-G color information. The lack of R-G color information therefore seems to influence human eye-movements.

The last image feature we look at is saturation. In those conditions in which the R-G channel is reduced, the effect sizes for saturation are significantly smaller than in the other two conditions (Figure 3.4, lower panel). Interestingly saturation is even higher in images devoid of B-Y color information than in naturally colored images.

The reduced saturation and virtual color contrast hint in the direction that we tend to look at different locations in R-G reduced images. Therefore we now analyze the similarity of eye-movements between different observers and different conditions.

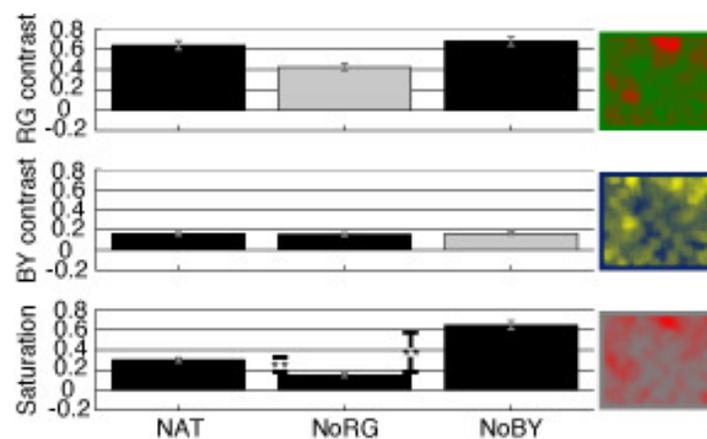


Figure 3.4 Color features.

Effect sizes (with SEM) for features R-G contrast, B-Y contrast, saturation in all three conditions of images.

### 3.2.3.2 Analysis of fixation distributions

The Reduction experiment is designed to examine the influence of color-channels on overt attention. If reducing one color channel alters human fixation behavior, then this channel contains salient information. We examined human fixation distributions in two ways. First, we analyze the congruency between different observers looking at the same image (same condition). A high congruency between observers is an indicator, that certain aspects of an image strongly attract attention. Second, we analyze the congruency of fixation locations between different conditions of the same image for each subject. This allows us to examine whether subjects tend to look at the same image regions irrespective of the type of stimulus modification. Taken together these measures allow us to assess, which aspect of color makes subjects look at shared locations, i.e. are salient.

The mean KL-divergence between different observers is 16.65, 17.8, and 16.56 bits for Natural, NoRG, and NoBY, respectively (Figure 3.5A). In NoRG the values are significantly higher than in both other conditions ( $p < 0.01$ ; two sided KS-test with Bonferroni correction). Higher KL-divergence values indicate less congruency between observers. This means that when R-G color information is absent, subjects tend to look at more dissimilar locations.

The mean KL-divergence between conditions is 17.31 bits for the comparison between Natural and NoRG and 16.92 for the comparison of NoBY and NoRG. The divergence for both conditions, which contain R-G color information, is much smaller, 14.95 bits (Figure 3.5B). The fixation locations of each subject on the same image are significantly less similar if R-G color information is missing. Only B-Y color information doesn't make subjects look at the same spots on the same image.

R-G color information therefore is not only salient when analyzing the population of subjects, but also when we look at each subject. These results suggest that R-G color information causally attracts overt attention in Rainforest images.

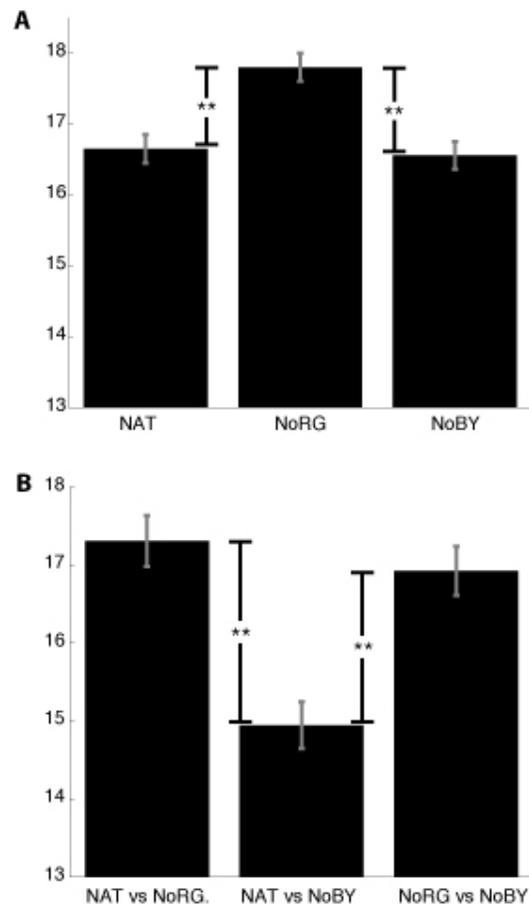


Figure 3.5 Congruency between fixation locations of color-normal subjects.

A: Mean KL-divergence (with SEM) between observers for NAT, NoRG, and NoBY. The divergence is significantly higher for NoRG compared to the other two conditions. In this condition the subjects' fixation patterns are more dissimilar. B: Mean KL-divergence (with SEM) between conditions.

### 3.2.4 Discussion

The results from Experiment 1 support the hypothesis of a causal influence of R-G color information on visual attention. The effect size for R-G color contrast in natural images is significantly higher than for B-Y contrast. If we completely remove all R-G color information from the images, the congruency between fixation locations of different subjects drops significantly compared to the natural baseline. In addition, subjects look at locations with significantly decreased saturation in the modified images. These strong effects can only be seen in images devoid of R-G color information but not in any other condition. The R-G color channel therefore is important for detecting “interesting” locations in Rainforest images.

## 3.3 Experiment 2

### 3.3.1 Introduction

The second stimulus manipulation which is able to uncover a causal relationship between color features and overt attention is the global manipulation of color information. In this experiment the manipulation consists of parametrically rotating the hue of each pixel in its isoluminant plane. The parametric rotation of hues allows us to state two extreme hypotheses about the influence of color contrasts on overt visual attention. If a color contrast has a causal influence on overt attention then it should have a constantly high effect size irrespective of color rotation. However, if other features are salient or higher cognitive processes involved, then the effect sizes should depend on the rotation angle. In the second case, we can assume that subjects look at the same image locations, irrespective of condition. This would yield effect sizes following a cosine when plotted against color-rotation angle. If we find this second pattern of results, we have to distinguish whether other features (bottom-up) or higher cognitive processes (top-down) are influencing overt attention.

### 3.3.2 Methods

#### 3.3.2.1 Stimuli

In Experiment 2 we used 84 images. The images were presented to the subjects in 12 different conditions. In each condition, the color of each pixel was rotated by a certain angle around the luminance axis. This caused changes in hue, for example a pixel that was red in the natural image would appear in a purplish blue if rotated by 90° clockwise, bluish green if rotated by 180° and greenish yellow if rotated by 270°. Rotation angles were chosen with exponentially increasing angles between 0° and 90°. Intermediate values were calculated according to:

$$r_n = 10^{\frac{n \cdot \log(90)}{6}}$$

and rounded to the next smaller integer. The first value ( $r_1=2^\circ$ ) was omitted because no visible difference to the 0° condition could be detected. Since images were rotated both clockwise and counterclockwise this led to the following twelve conditions: 0°, 4°, 9°, 20°, 42°, 90°, 180°, 270°, 318°, 340°, 351°, 356°.

After these modifications, the stimuli were converted to RGB images for presentation on the screen. A complete sequence of all rotations is reproduced in Figure 3.6.

Subjects were seated 80 cm from the monitor surface, which yielded approximately 28×21 degrees of visual angle for our stimuli.

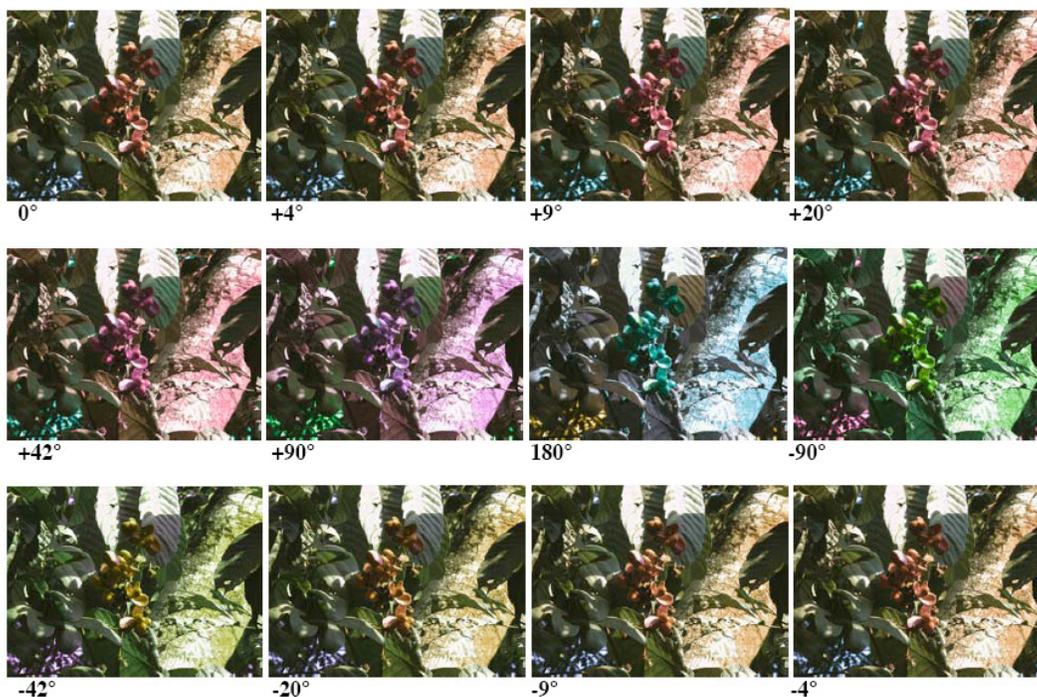


Figure 3.6 Experimental conditions in Experiment 2.

We depict the same image in all 12 different conditions of color rotation.

### 3.3.2.2 Experimental Design

For the Rotation experiment we used a between subject design. Every subject saw all 84 images, each in only one condition. Since we had 12 conditions, subjects saw 7 images for each condition. The order of conditions and images was randomized between subjects. On average, each image was seen 3 times in each condition by all 36 subjects taken together. The 84 trials were divided into 2 blocks of 42 trials, and subjects could take a break in between blocks.

### 3.3.2.3 Image statistics

One critical assumption of bottom-up driven saliency is that singular peaks in one feature channel should contribute to saliency more than multiple peaks in another channel. In order to extend our analysis of color features we therefore assessed the peakiness of the different feature distributions. We defined the following measure of peakiness: we summed all feature values that are bigger than the sum of mean and two standard deviations. This sum of very high feature values was then divided by the number of image pixels taken into account.

### 3.3.3 Results

#### 3.3.3.1 Analysis of image features

##### 3.3.3.1.1 Hue

A color can be described by its hue, saturation, and lightness. It's conceivable that a particular hue like e.g. a bright red attracts overt attention. However, in analogy to luminance features examined in earlier studies, we expect only weak influence of hue on overt attention and rather strong effects of color contrast. Since there is no clear way to calculate effect sizes for hue, we examined the distribution of hues at fixated locations. If hue causally influences overt attention then we should find always the same hue at fixation locations, irrespective of color rotation.

Our analysis reveals that the distribution of hues at fixated locations changes systematically with color rotation (Figure 3.7). Looking e.g. at the peaks of the distributions we can see a shift almost exactly coincident with the degree of color rotation. The peak in the natural condition is at about  $115^\circ$ , while it is at  $25^\circ$  in the condition, in which hues are rotated by  $90^\circ$ . This indicates that human overt attention is not attracted by particular hues.

The shape of the hue distributions seems very similar in all conditions. In order to analyze quantitatively the similarity between the hue distributions of the different conditions we applied cross-correlation analysis. This allows us to verify that the shift of distributions in fact matches the color rotation of images. Cross-correlation was almost always maximal when the shift between the hue distributions corresponded to the relative color rotation between the two conditions the distributions came from (Figure 3.8A). Only in some comparisons the maximum was shifted by one degree. The correlation coefficient between relative rotation angle and the shift that yielded maximum cross-correlation is  $r = 1.0$ . The maximum cross-correlation values between different hue distributions are very close to 1 (Figure 3.8B). This indicates that all hue distributions are almost identical.

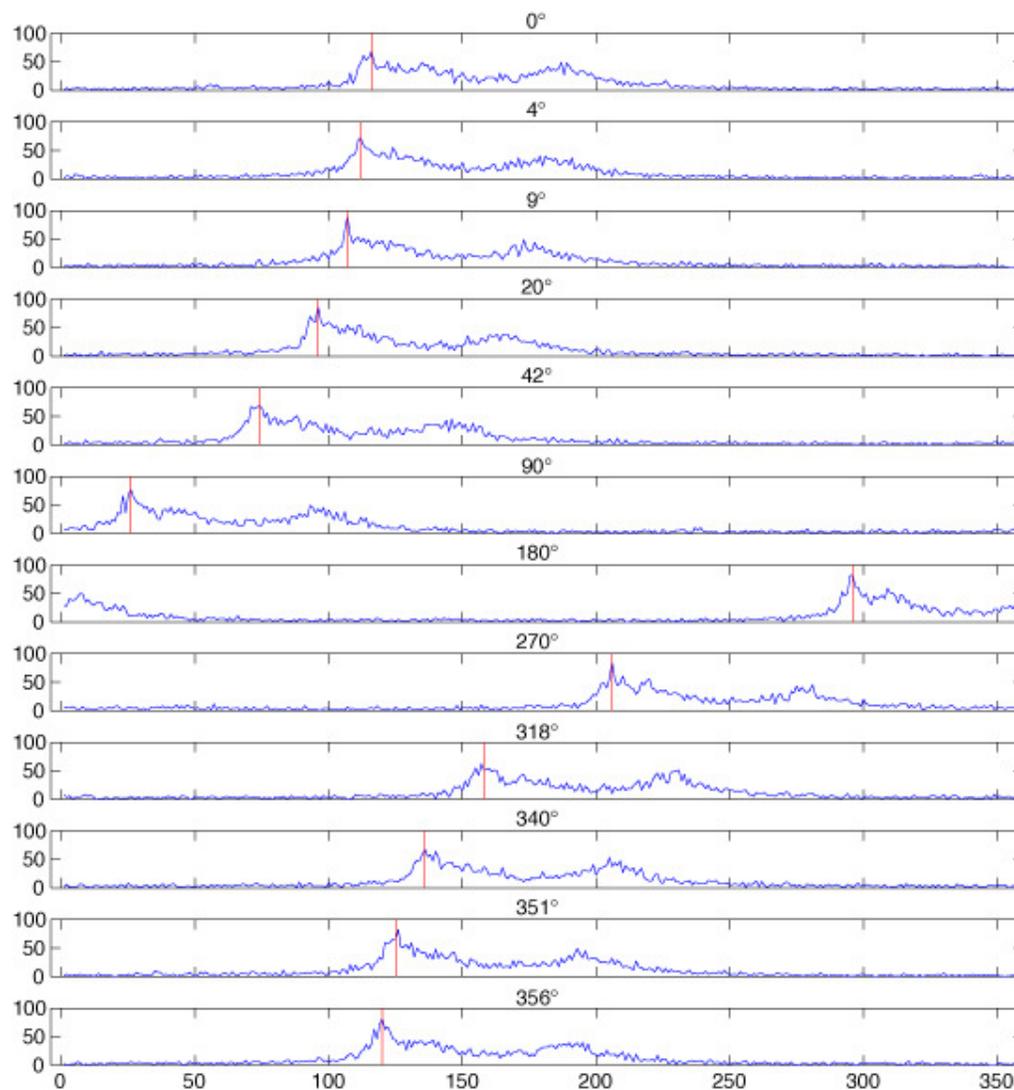


Figure 3.7 Influence of hue on overt attention.

Histograms of mean hue (in degrees) in a 21x21 pixel patch around fixation locations for all conditions. The red line indicates the location of the maximum value in the first graph (no rotation) and is drawn in the other graphs to indicate where this maximum would be expected due to the color rotation.

In conclusion, we show that hue is not influencing overt visual attention. The high similarity between hue distributions of different conditions and size of the shift suggests that subjects look at almost the same image locations irrespective of color rotation.

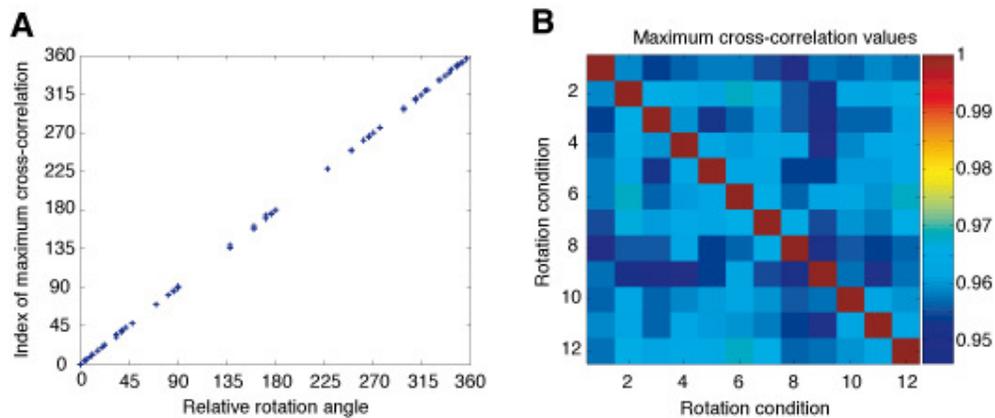


Figure 3.8 Influence of hue on overt attention.

A: Shift at which the cross-correlation between two hue distributions is maximal plotted against the relative rotation angle between the conditions the hue distributions were taken from. B: The maximum cross-correlation values for all combinations of conditions.

### 3.3.3.1.2 Color contrast

The parametric rotation of hues makes it possible to state two extreme hypotheses about the influence of color contrasts on overt visual attention. If a color contrast has a causal influence then it should have a constantly high effect size irrespective of color rotation. However, if there are other (probably top-down guided) features, which are really salient, then the effect sizes depend on the rotation angle. In case these salient features do not depend on color or if subjects are able to compensate for color rotation, we can assume that subjects look at the same image locations, irrespective of condition. This would yield effect sizes following a cosine when plotted against rotation angle. The two cosines for R-G and B-Y contrast would always go in the opposite direction.

In naturally colored images ( $0^\circ$  rotation) R-G contrast is more salient than B-Y contrast. This coincides with the findings of our earlier study. With increasing rotation angle the R-G contrast first slightly increases, then reaches a minimum at  $90^\circ$  and a local maximum at  $180^\circ$ . B-Y contrast shows the inverse pattern (Figure 3.9). Analysis on  $41 \times 41$  and  $121 \times 121$  pixel patches showed slight variations in the absolute effect sizes (higher for smaller patches, lower for larger patches), but the global pattern remains completely unchanged. The obtained effect sizes almost perfectly match the second hypothesis.

This means that no color contrast along a specific axis in color space influences overt visual attention causally in Rainforest images. It is conceivable, that color contrast is evaluated in a hue invariant way. Even though the images with large color rotations look very strange, one can still detect easily all the structural elements of the image.

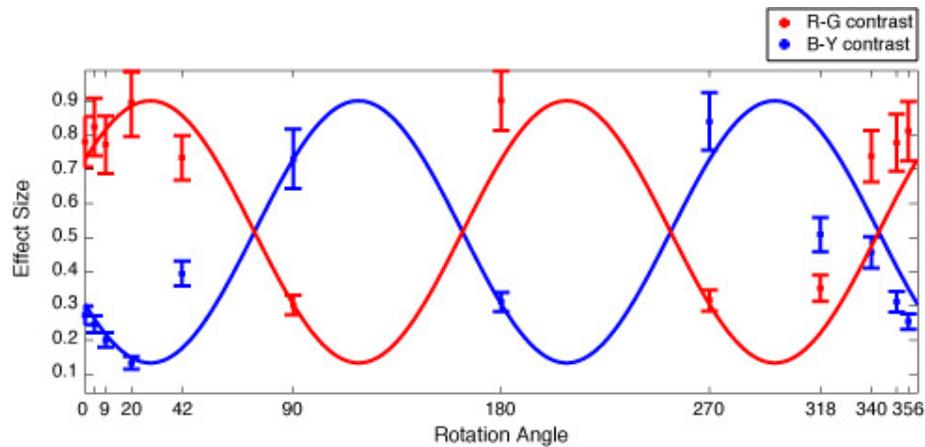


Figure 3.9 Influence of color-contrasts on overt attention.

Mean effect sizes (with SEM) for R-G (red marks) and B-Y (blue marks) contrast, calculated on 81x81 pixel patches, in the different color rotation conditions. The cosines reflect the hypothesis that the same image locations are fixated in all conditions, irrespective of its color contrast. Their maximum is fitted to the largest effect size found, and the amplitude is the difference between the largest and the smallest effect size. The cosines are shifted by  $34.5^\circ$  to the right to correct for the asymmetry in image statistics described in the Asymmetry section below. The labeling of tick-marks at the  $4^\circ$  and  $351^\circ$  condition are omitted for better readability.

Here we have to consider the influence of image statistics of the color features on fixation behavior. It is possible, that subjects are attracted by the feature with the highest variations in the image. These image statistics change along with the parametric rotation of the colors. Therefore we examined the peakiness of the color-contrasts in all images and conditions. And indeed we find the pattern of peakiness matches the pattern of the color contrast effect sizes very well (Figure 3.10). For both color contrasts the correlation between mean effect size and mean peakiness over conditions is  $r = 0.93$ .

This indicates that higher peaks in the color feature maps make a feature more salient. This would constitute a mechanism that is not tied to a specific axis in color space, but interacts with global image statistics. However since this is only a correlational observation, we have to be careful interpreting this result and present a more detailed analysis below.

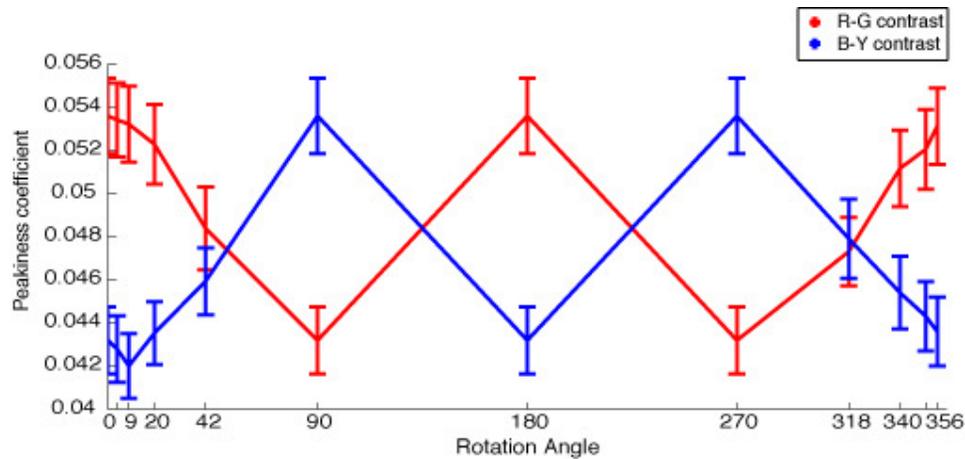


Figure 3.10 Peakiness of color contrasts.

Mean peakiness values (with SEM) for R-G (red line) and B-Y (blue line) color contrasts, calculated for the same patch size as Figure 3.9. There is a correlation of 0.93 between the mean peakiness and mean effect sizes for both color contrasts.

### 3.3.3.1.3 Asymmetry

In Figure 3.9 we can see that the R-G contrast effect sizes are markedly higher in 20° and 42° rotation, than in the 340° and 318° rotation. The inverse holds for B-Y contrast. This means that we have an asymmetry of effect size between rotating clockwise or counter-clockwise. We further elucidate this asymmetry by investigating statistical properties of our images. The distribution of colors in patches containing fruit has its greatest variance along an axis that is 34.5° below the R-G axis. This angle obtained by principal component analysis is almost identical to the value obtained by a linear regression. Color rotation in the counter-clockwise direction causes this axis to move closer to the R-G axis, thereby increasing the red-green contrast. Rotation in the clockwise direction has the opposite effect (Figure 3.11). Therefore, the asymmetry effect can be explained by image statistics of frequently fixated image patches.

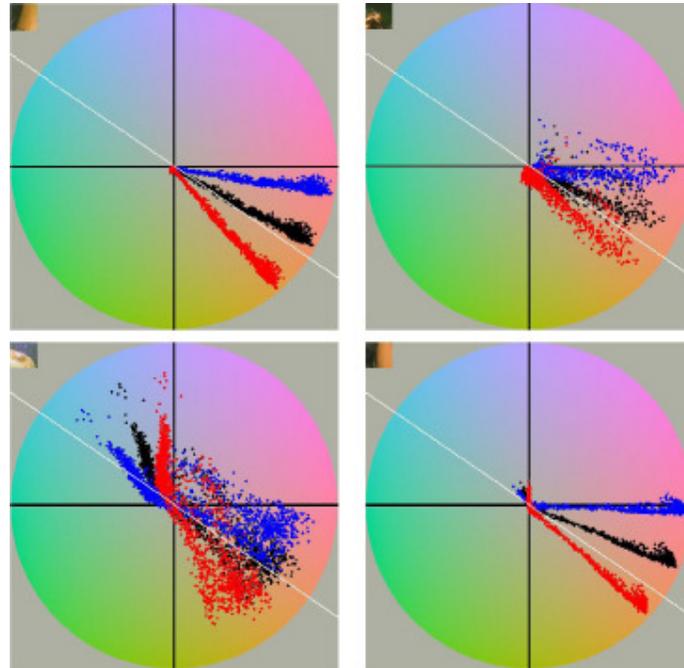


Figure 3.11 Analysis of the asymmetry effect.

The black dots indicate the colors of all pixels in the image patch shown in the upper left corner of each figure. All colors are projected onto the isoluminant plane through the origin of the DKL color space. Blue dots represent the same pixels for 20° counter-clockwise (+20°) rotation, while red dots represent the 340° rotation. The white line indicates the angle of the first PCA component on all patches around fruit fixations.

#### 3.3.3.1.4 Saturation

The last color feature we analyze in the Rotation experiment is saturation. For saturation we can take one hypothesis from the analysis of color-contrast. If we find constantly high effect sizes for saturation in all conditions, then this feature should be salient. Since this feature was held constant at each pixel throughout all conditions, different subjects fixating exactly the same locations in one image would lead to exactly the same effect sizes for all conditions. Saturation at fixated locations is about 50% higher than at control locations (Figure 3.12). However, there are three conditions, in which the effect sizes are strongly different from 0.5. These conditions are 9°, 42°, and 318°. The latter two are also those conditions we find the biggest asymmetry for color-contrasts. But since saturation is held constant at each pixel, this effect, unlike the effects for color-contrast, cannot be explained by image statistics.

Since the saturation effect is positive and relatively constant across conditions, we can assume that this feature is either salient or correlated to features selected by higher-order mechanisms.

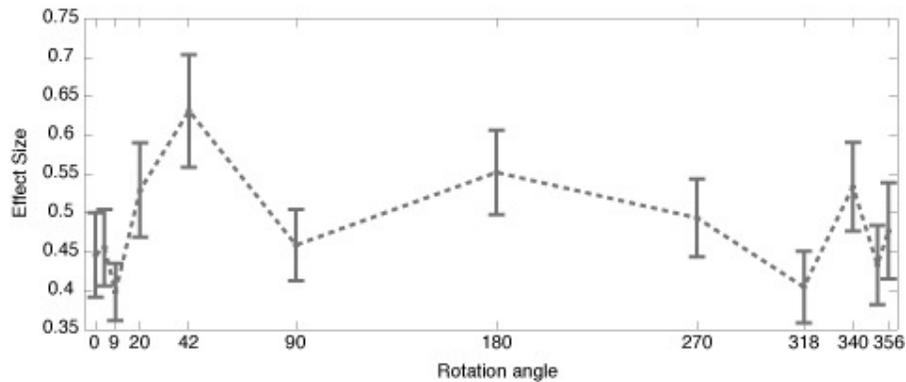


Figure 3.12 Influence of saturation on overt attention.

Effect size (with SEM) for saturation in all different experimental conditions.

### 3.3.3.2 Analysis of fixation distributions

The results for the color contrasts let us that subjects fixated the same locations no matter how the images were color-rotated. Since we showed each image only once to each subject we cannot compare the fixation locations of the same subject between different conditions.

However, we can do an inter-observer congruency analysis comparable to the Experiment 1. It is not exactly identical, since the different subjects didn't always see the same condition of images either. However, if the KL-divergence values for inter-observer congruency are the same as in Experiment 1, this would mean that the global color rotation does not influence attentional selection of fixation points.

Mean KL-divergence between subjects (which saw the image mostly in different conditions) is 15.54 bits. This value is lower than the divergence between observers in the Color Channels experiment. This indicates that viewing behavior is rather constant when stimuli are rotated in color space and that when color information is reduced (Color Channels experiment, see above) viewing behavior is changes drastically.

This suggests that subjects are able to compensate for the global change in hue. They most probably fixate the same locations in color-rotated and naturally colored images.

### 3.3.4 Discussion

On first sight the results from Experiment 2 are not in line with the hypothesis of a causal influence of color features along specific axes in color space. If R-G contrast causally attracted gaze, then it should do so, no matter whether or not the colors in the image are rotated. Violating this prediction we find a strong dependency of both color contrast's effect sizes on rotation angle. The same holds for the hue at fixated locations. In contrary, saturation effect sizes are not affected by color rotation. However, due to the modification algorithm

used to generate our stimuli, this feature is identical for each pixel between conditions. Therefore, our pattern of results is generated by subjects looking at approximately the same image locations irrespective of the rotation angle. Support for this notion comes from the analysis of inter-observer congruency. Even though subjects saw the images mainly in different conditions, the congruency between observers was in the range of the natural images from Experiment 1.

These observations are compatible with a hue invariant evaluation of saliency. This calls for an interaction of different color channels. Several brain areas like primary visual cortex (V1; Li, 2002), visual areas V4 and IT (Mazer & Gallant, 2003) have been postulated to be involved in computation of saliency and contain color selective cells. Response properties and connectivity of V1 neurons could be suited for the computation of a saliency map (Li, 2002). In order to explain the results of the color-rotated images, this processing needs to encompass a hue invariant peakiness measure. In V1 there are 3 times as many cytochrome oxidase-rich blobs coding R-G as for B-Y color information (Ts'o & Gilbert, 1988). It is unclear whether this would determine an advantage for the R-G system to influence overt attention. Furthermore, it is not obvious that a separation of neurons with different color preferences in separate blobs supports hue invariant saliency computation. This does not make V1 a promising candidate for the computation of color saliency.

Another hint for the processing of invariance mechanism is color constancy. The color-rotations can be regarded as a change of the scene's illumination. The visual system of humans is able to discount such illumination changes to a certain degree. One brain area suggested to compute color constancy is V4 (Zeki, 1983). However it is very likely not the only brain area involved in this process (Schein & Desimone, 1990). A study on patients with circumscribed brain lesions showed that middle temporal areas are crucial for color constancy (Rüttiger, Brain, Gegenfurtner, Petersen, Schönle, & Sharpe, 1999). Therefore it is also plausible that higher visual areas in the ventral stream compute color constancy and contribute color information to a saliency map of visual attention.

## **3.4 Experiment 3**

### **3.4.1 Introduction**

In many areas of neuroscience, the examination of persons with certain inabilities has led to new insights about brain function. Therefore we conducted an experiment, in which we showed selected conditions of images to color-blind (deuteranope) subjects. Due to altered sensitivities of cones in the retina these subjects have difficulties in discriminating between reddish and greenish hues. Since the genetic alteration in deuteranopes is inherited these subjects experienced the reduced visual information all their life. For this experiment there are two extreme outcomes conceivable. Either a deficiency in color processing alters attentional processing dramatically or there are compensating mechanisms. In the former case a deficiency in color processing should influence overt visual attention in an observable way. This means that we should be able to find distinctive differences between the deuteranope and the normal-sighted subjects with respect to stimulus features and fixation distributions. In case there are compensating mechanisms for the missing color information, we try to examine whether they are based on other low-level features or higher processes.

### **3.4.2 Methods**

#### **3.4.2.1 Subjects**

12 subjects responding to an advertisement took part in the experiment 3. We administered the Ishihara plates (Kanehara Trading, Tokyo) and the Farnsworth-Munsell 100-Hue test under daylight conditions. Based on the scores and error patterns in both tests, we classified 8 of these subjects as deuteranope. We will only report the results for these subjects.

#### **3.4.2.2 Stimuli**

In Experiment 3, we employed the same 40 images as in Experiment 1. To increase the probability of assessing differences between deuteranope and normal color-sighted subjects, we presented these images in conditions with strong effects in Experiment 1 and 2. Therefore we employed the conditions NoRG and NoBY, as well as 90° as well as 270° color-rotated. To obtain a baseline for effects without any color information available, we also presented these images in grayscale.

Subjects were seated 80 cm from the monitor surface, which yielded approximately 28×21 degrees of visual angle for our stimuli.

### 3.4.2.3 Experimental Design

The experiment was conducted in a session with 2 blocks of 120 images. The order of presentation was randomized for each subject, with the constraint that no image was presented twice in a row.

## 3.4.3 Results

### 3.4.3.1 Analysis of image features

Deuteranope subjects have problems in discriminating yellowish red from green based on their hue. These hues however are very prominent among the rainforest pictures we use in this study, Therefore we expect that color features influence deuteranope subjects different than normal sighted subjects.

To our surprise, the effect sizes of the color-blind subjects are almost identical to those of normal subjects for color-rotated images (Figure 3.13). The effect size for R-G contrast is significantly higher than for B-Y contrast in naturally colored images. Deuteranope subjects tend to look at locations with significantly higher B-Y contrast in color-rotated images compared to Natural. The B-Y contrast effect size is 0.7 for 90° rotated and 0.68 for 270° rotated respectively, while it is only 0.2 in Natural. As for the normal subjects, we find the opposite pattern for R-G contrast.

For the color-reduced stimuli, the effect sizes between the two populations of subjects are very similar, too. The only difference we find is with the virtual color contrast. While the difference between the effect size for R-G contrast in Natural images and virtual contrast in NoRG images is only 0.12 for color-blind subjects, this difference is 0.22 in normal sighted subjects. This indicates that even in the absence of R-G color information deuteranope subjects are more likely to look at locations that would have a high R-G contrast. It is a first hint that color blind subjects employ compensatory mechanisms and therefore in the absence of R-G information perform better than normal subjects.

In order to refine our approach of averaging over all fixations in a trial we analyze the data time resolved at the level of single fixations. Since fruits in front of foliage have a high R-G contrast, we exclude for this analysis all images with fruits within 2 degrees of the initial fixation cross. This leaves us with 22 images.

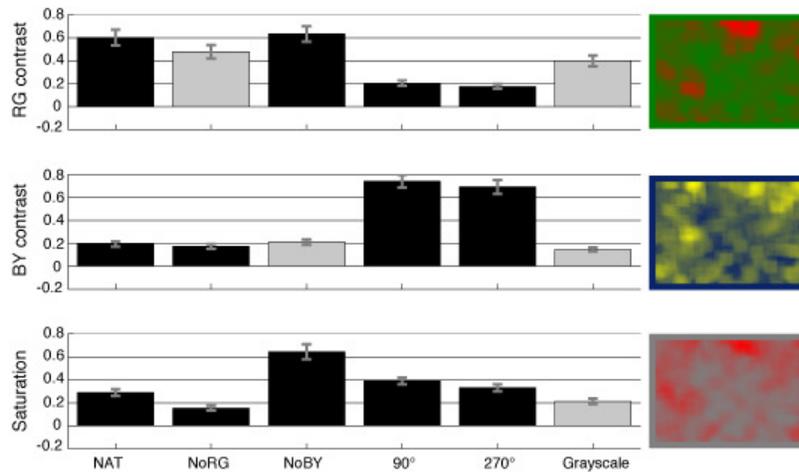


Figure 3.13 Color features in deuteranope observers.

Effect sizes (with SEM) for features R-G contrast, B-Y contrast, and saturation in all conditions of images in deuteranope subjects.

For the color-normal subjects from Experiment 1 the maximum peak of R-G contrast is reached within 2 fixations (Figure 3.14). After this fixation the R-G contrast values decline continuously. In contrast, the deuteranope subjects first look at patches with lower R-G contrast. They need 4 fixations to reach the patches with maximal R-G contrast effect size. This means that there is a delay of 2 fixations that deuteranope subjects need to reach maximum R-G contrast. We find no such delay for the other color features, B-Y contrast and saturation.

Summing up, we do not find differences in the averaged color feature effect sizes between color-normal and dichromatic subjects. The dichromatic subjects are able to compensate for their deficit in color discrimination. However, when examining the effect sizes on the basis of single fixations we find that dichromatic subjects reach the maximum of R-G contrast effect size 2 fixations later than color-normal subjects. This is an indication for a top-down compensation process.

### 3.4.3.2 Fixations on Fruit

One prominent hypothesis for the emergence of trichromatic vision in old world primates states that trichromacy helped to discriminate ripe fruit from foliage, thus increasing the efficiency of foraging. Therefore we would expect to find more fixations on fruits in color-normal subjects. In images devoid of R-G color information we expect similar percentages in both groups of subjects, because trichromatic subjects are missing important color information.

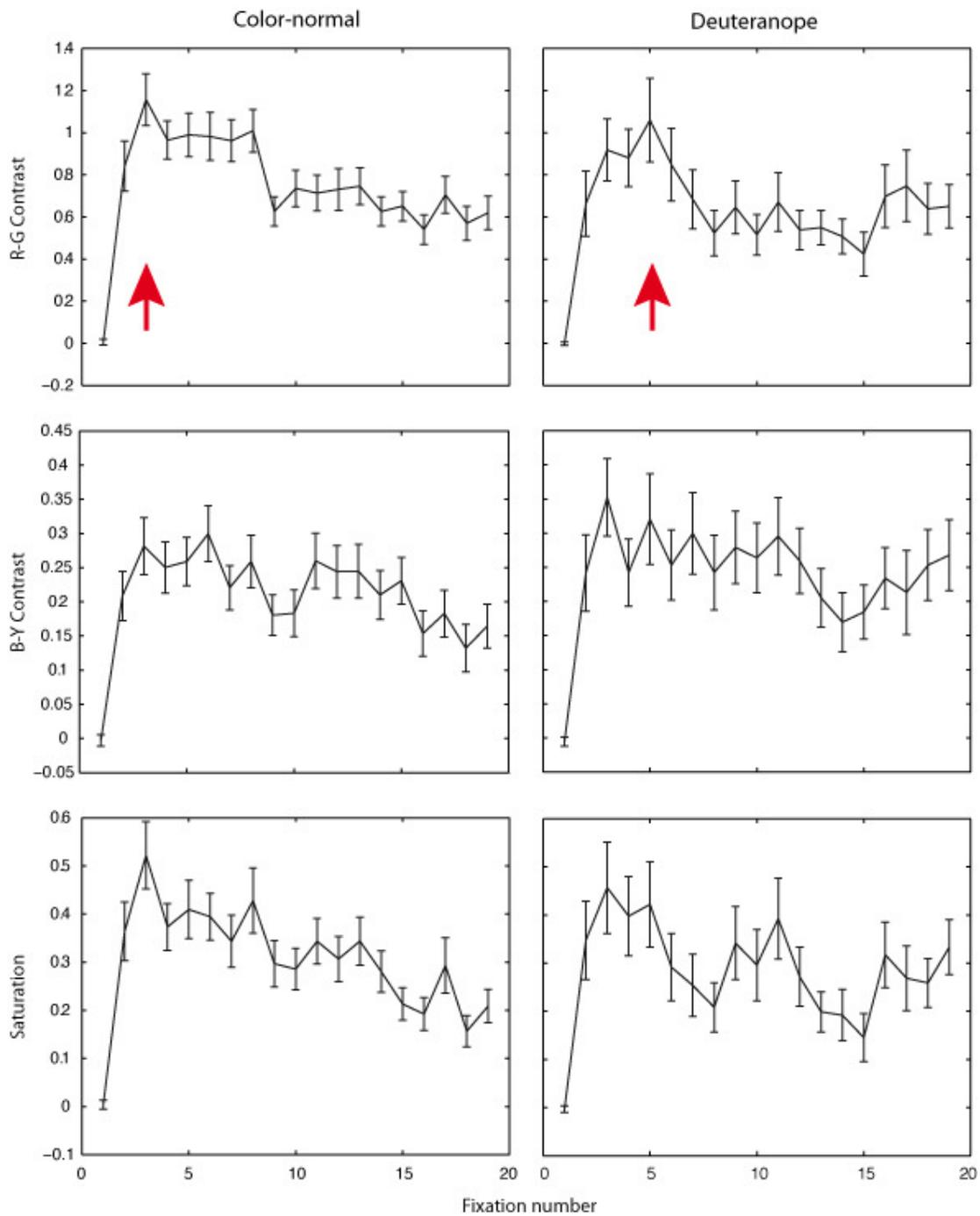


Figure 3.14 Color features at each fixation.

Effect sizes (with SEM) for the features R-G contrast, B-Y contrast, and saturation. The maximum value for R-G contrast is reached within 2 fixations for the color-normal subjects, while deuteranope generally reach the maximum value within 4 fixations. We do not find a comparable difference in other color features.

The average percentage of fixations on fruit in all naturally colored images is similar between color-normal (25%) and deuteranope (26.4%) subjects. In images devoid of R-G color information the percentages are 21.6% and 26.7%, respectively. For color-normal subjects, we

find the expected slight drop in fruit fixations in images devoid of R-G color information. For deuteranope subjects, however, the percentage of fixations on fruits remains constantly high. This result contradicts our initial hypothesis that deuteranope subjects would be less efficient in detecting fruits and therefore fixate less. It indicates that deuteranope subjects are less influenced by the absence of R-G color information than color-normal subjects.

We do not find the expected advantage of color-normal subjects in fixating fruit objects in all Rainforest images. The color-normal subjects show the expected dependency of percentage of fruit fixations on the availability of R-G color information. However, the deuteranope subjects look at fruits even slightly more than trichromatic subjects in naturally colored images and show no reduction in fruit fixations with missing R-G color information. This shows that deuteranope subjects compensate for the missing R-G color information.

Another important question concerning fixations on fruits concerns the influence of color features. We therefore analyze the effect size of different color features for single fixations made on fruits. This reveals how much and when color features of fruits influence gaze.

The effect size of R-G contrast of fixations on fruits shows exactly the same pattern as in Figure 3.14, in which we analyzed all fixations. For deuteranope subjects we see the peak of effect size for the 5<sup>th</sup> fixation, while it is 2 fixations earlier in color-normal subjects (Figure 3.15).

Color-normal subjects detect these high R-G contrast patches on fruits very fast and fixate them. In contrast, deuteranope subjects first look at parts of fruits, which have a lower R-G contrast. This indicates that fixations on fruit objects can explain the results for the detailed analysis of all single fixations.

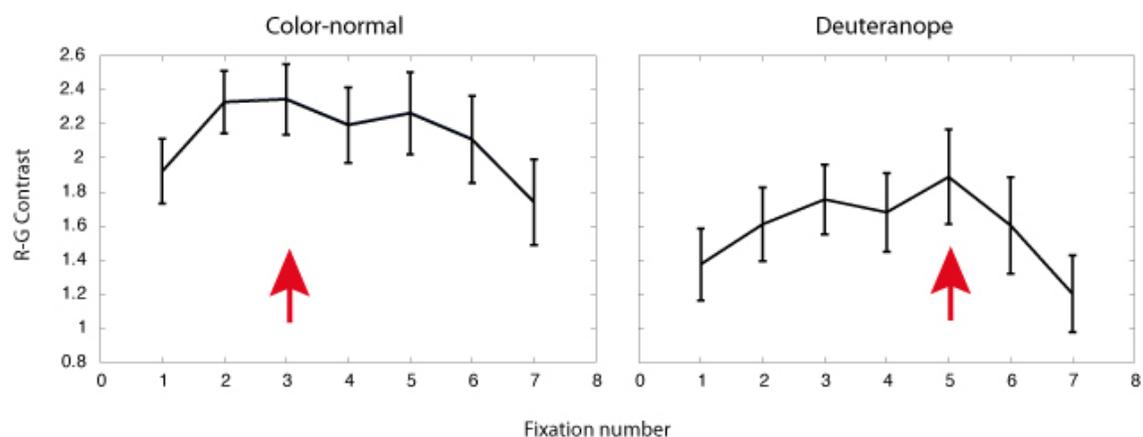


Figure 3.15 R-G contrast at each fixation on a fruit object.

The maximum value for R-G contrast is reached within 2 fixations for the color-normal subjects, while deuteranope generally reach the maximum value within 4 fixations. We do not find a comparable difference in other color features.

### 3.4.3.3 Analysis of fixation distributions

How does the different apperception of the visual environment influence the eye-movements of deuteranope subjects? The color deficiency could impair the perception of the leaves and fruits depicted in the scenes. Therefore it is possible that the fixation behavior of deuteranope subjects differs from normal sighted subjects. We examined several aspects of fixation behavior. To make reading easy we kept the order of description of the previous experiments. We first report the results on congruency of fixation locations between the different deuteranope subjects, as well as the congruency between different conditions. Then we describe the similarity of fixation locations between deuteranope and normal sighted subjects. In general, the KL-divergence values are significantly smaller than those of the color-normal subjects for all comparisons. Deuteranope subjects exhibit a significantly higher inter-observer congruency than the normal sighted subjects of the Channels experiment, who saw the same images. As for the normal sighted subjects, inter-observer congruency is lowest (highest KL-divergence) in images devoid of R-G color information and grayscale images (Figure 3.16A). Looking at the congruency between conditions, we find a difference to the color-normal subjects. The congruency of fixation locations of deuteranope subjects does not differ significantly between images with or without R-G color information (Figure 3.16B). We find slight differences in the congruency of fixation location analysis between color-normal and color-blind subjects. The color-blind subjects exhibit a higher inter-observer congruency and a higher congruency of fixations between conditions than color-normal subjects. The overall pattern for inter-observer congruency is very similar to the one of color-normal subjects. But the fixation congruency between conditions in each individual differs from the one in color-normal subjects. There is not a significantly lower congruency in fixation locations if we compare images with and without R-G color information. In this measure we compare the fixation locations of the same subject in different conditions. Therefore this result could be an indication for individual differences between deuteranope subjects. It has to be noted, however, that the number of participants is smaller than for normal subjects and the variation between them somewhat larger. These two aspects limit the statistical power of our analysis.

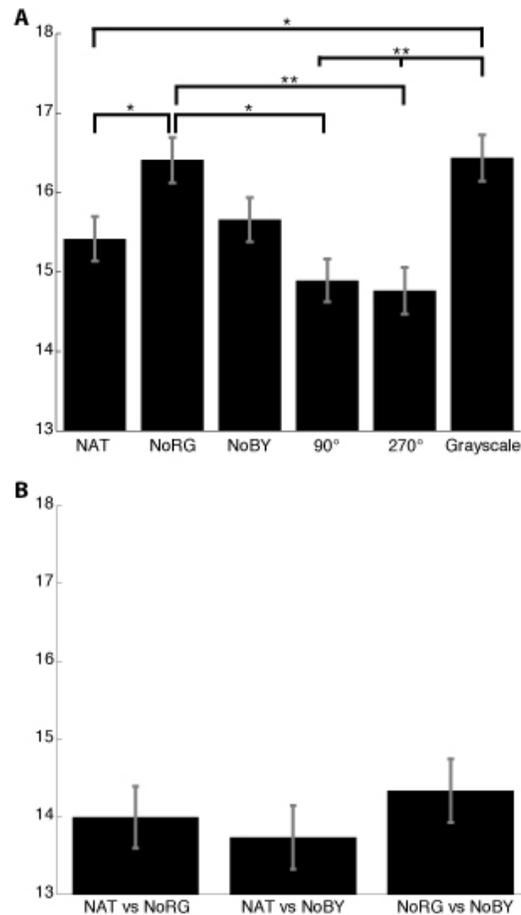


Figure 3.16 Congruency between fixation locations in deuteranope observers.

A: Mean KL-divergence (with SEM) between different deuteranope observers for NAT, NoRG, NoBY, 90 and 270 rotated, and Grayscale. The divergence for conditions devoid of R-G color information is significantly bigger than in all other conditions (except NoBY). B: Mean KL-divergence (with SEM) between conditions.

### 3.4.4 Discussion

The analysis of eye-movements in deuteranope subjects revealed several similarities with the color-normal subjects. This can be taken evidence for a high-level compensation process for the missing color-information. Further evidence for a compensation process in deuteranope subjects comes from the analysis of fixations on fruits. In a recent study we showed that fruits have a significantly higher R-G contrast than other fixated locations in Rainforest images (Frey, Wirz, Willenbockel, Troscianko & König, 2006). While color-normal subject show a drop in fixation percentage of fruit with missing R-G color information, deuteranopes' percentage does not change.

However, there are also some differences. The nature of these differences allows us to draw conclusions on the nature of the compensation process. Deuteranope subjects reach maximum R-G contrast 2 fixations later than normal sighted subjects, which is about 500-600ms in our

data. This time course of R-G contrast in deuteranope subjects indicates that the compensation in deuteranope subjects stems from a top-down process.

A likely explanation for this compensation process is that deuteranope subjects detect certain types of fruit objects based on other features than R-G contrast or scene composition. They then use prior knowledge to fixate on those patches of fruits that have a high R-G contrast.

### ***3.5 Discussion of Experiments 1-3***

In the present study, we expand previous analyses of the salience of color features by parametrically changing image properties. This allows us to determine whether there is a correlative or causal influence of these features on overt attention. Removing the available R-G color information in Rainforest images significantly alters the subjects' fixation behavior. This clearly speaks in favor of a causal stimulus dependent (bottom-up) influence of the R-G color channel on overt attention. However, if we globally rotate colors in these images, there is no causal relationship between specifically the R-G contrast and overt attention. R-G and B-Y contrast alternate in saliency with rotation angle. This is an indication for an invariance mechanism, which guides attention irrespective of specific axes of color-channels. Comparing effect sizes of color features between color-normal and deuteranope subjects at the level of single fixations reveals that R-G contrast attracts overt attention earlier in normal-sighted than in deuteranope subjects. Therefore we assume a top-down compensation process in deuteranope subjects. This indicates different levels of attentional control influence overt visual attention in natural images.

#### **3.5.1 Different levels of attentional control**

For color-normal subjects attentional control is very likely to be influenced by visual areas in the ventral stream. Li (2002) proposed a pre-attentive computational mechanism for the generation of a saliency map in V1. However, it is unclear in how far the known properties of V1 are compatible with the results on the color-rotated scenes as discussed above. Therefore it seems more likely that higher level areas of the ventral stream involved in color processing like V4 and IT are contributing to a saliency map of colored scenes. Importantly, these are also implicated in color-constancy which is compatible with a hue invariant computation of saliency. Further evidence for a fast parallel attentional mechanism comes from neurophysiological experiments in macaque V4 (Bichot, Rossi, & Desimone, 2005) and IT (Mazer & Gallant, 2003). It is, however, possible that several levels of the visual hierarchy contribute to a computation of saliency.

In deuteranope subjects we find evidence for a top-down driven compensation process. The lack of processing color information along the R-G axis is compensated by higher level concepts at the cost of a significant delay. Recent studies propose that top-down control of visual attention is mediated by superior frontal and intraparietal cortices (Corbetta and Shulman, 2002). In the light of these results the present observations show that different levels of the visual hierarchy are involved in the control of overt visual attention in natural scenes.

### 3.5.2 Other image features

This study was specifically designed to address the influence of color features on overt visual attention. There are several other luminance-defined stimulus features, which have been studied in recent experiments. We will shortly discuss the findings of these experiments and their relation to the present study.

One of the first image features proposed to influence overt attention was luminance contrast. Several studies showed elevated luminance contrast at fixated compared to non-fixated regions (Reinagel and Zador, 1999; Parkhurst and Niebur, 2003). However, Einhäuser and König (2003) could show that when locally altering luminance contrast there was no causal relation. In addition, two studies of our lab using the physiologically plausible DKL color space showed just weak or missing salience of luminance contrast in different categories (Frey, König, and Einhäuser, 2007; Frey, Honey, and König, 2008). Therefore luminance contrast most likely isn't a feature to compensate for altered color information.

The second-order contrast in the luminance channel is another possible candidate. Using a modified model of bottom-up attention, which incorporated this so-called texture contrast, Parkhurst and Niebur (2004) could replicate in a model the experimental results of Einhäuser and König (2003). However, we found texture contrast to be salient only in images of Landscape and Man-Made objects (Frey, König, and Einhäuser, 2007; Frey, Honey, and König, 2008). Therefore the salience of this feature is definitely limited to certain categories. Other luminance defined image features that could be possible candidates are edge density or more complex line elements. Mannan, Ruddock, and Wooding (1996) analyzed several image features including luminance-defined edges, but could show that they do not really attract attention. More complex line elements like curved edges or junctions are more likely candidates. Using their definition of intrinsic dimensionality, Krieger, Rentschler, Hauske, Schill, & Zetzsche (1996) showed that the bispectra of image regions selected by complex line elements (i2D operators; Zetzsche and Barth, 1990) are similar to the bispectra of image regions fixated by human observers. The i2D operators are comparable to stimuli that drive

end-stopped cells found mainly in area V2 of visual cortex. Combining intrinsic dimensionality operators and self-information, Saal, Nortmann, Krüger, and König (2006) generated saliency maps that predicted human fixations very well in different categories of images. However, these operators are highly nonlinear and therefore require special techniques for modeling and analysis.

### **3.5.3 Timing of bottom-up and top-down**

Several visual search studies have addressed the timing of bottom-up and top-down attentional control. Stimulus-driven attention is believed to control attention at the very early time window Müller and Rabbit (1989) showed that valid peripheral cues presented around 175ms before the onset of a search display increased the probability of detecting the target. For valid central cues however, the stimulus-onset asynchrony needed to be about 400ms. These results suggest that top-down control elicited by the central cue occurs later than bottom-up control elicited by the peripheral cue. In addition, these two processes seemed to have a different time course of operation. A similar finding has been reported by van Zoest, Donk, and Theeuwes (2004). They required subjects to make a speeded saccade towards a target embedded in several non-targets. If a salient distracter was presented on a trial, the percentage of saccades to the target was low for fast saccades. Somewhat slower saccades with a latency of around 300ms were aimed significantly more often to the target. This indicates that bottom-up control operates very early in a trial and then top-down sets in. These results are in line with a timing account of visual attention (van Zoest, Donk, and Theeuwes, 2004). According to this perspective, visual selection may be either bottom-up or top-down driven as a product of response time.

However while in the visual search study of van Zoest et al (2004) already the first saccade was susceptible to top-down influence, we see the maximum peak of the bottom-up feature effect sizes only at the target of the second saccade. Since in our study the second saccade started around 630ms after stimulus onset, the peak bottom-up contribution is in the time-range of top-down control, as determined by visual search experiments. One possible explanation is definitely the complexity of the stimuli used. The Rainforest images used in this study typically contain several locations with very high values for color and luminance-defined features. It is very likely that subjects first make an exploratory saccade to the closest “interesting” location. This difference between the natural scenes and visual search experiments could probably be further examined by using increasingly more crowded search displays.

A previous eye-tracking study employing natural scenes examined the time course of several

image features (Tatler, Baddeley, and Gilchrist 2005). The only color feature analyzed was chromaticity, a feature defined by the combination of outputs from R-G and B-Y color-opponent channels. They found that this color feature was about equally salient over the time course of fixations. There was no increased salience for early fixations. There are two possible reasons, why their results differ on first sight from the results of the current study.

First, Tatler and colleagues used images of indoor and outdoor scenes in a city. Using similar stimuli we found R-G and B-Y contrast to be much less salient than in Rainforest images (Frey, Honey, and König, 2008). Therefore, the choice of stimuli should play a role for the difference between the two studies. Second, they examined a feature, which combined R-G and B-Y opponent color-channels. In images of man-made objects, which come close to what Tatler and colleagues used as stimuli, we found very low correlations between the color-contrasts in these two channels. Therefore it could be possible that this averaged chromaticity feature just is not sensitive enough.

### **3.5.4 Deuteranope subjects**

We have to be careful when trying to generalize the results from studies involving observers with color vision deficiencies. On the one hand, several studies revealed that dichromatic observers also exhibit other color deficits. For example, Regan, Reffin, and Mollon (1994) showed that protanope and deuteranope subjects have higher tritan thresholds than color-normal observers. This indicates that color-channels that were assumed to be unaffected in dichromats, also showed impairments. On the other hand, there seems to be considerable heterogeneity among deuteranomalous observers for the wavelength of peak sensitivity (Shevell and He, 1997). Since we did no extended psychophysical testing on a Nagel anomaloscope and no genetical sequencing of their opsin gene arrays, we can not classify our subjects as deuteranopes with absolute certainty. However, we tried to obtain a homogenous group by selecting only those subjects who had the highest deuteranope-score in the Ishihara test and exhibited the typical error pattern in the Farnsworth-Munsell 100-Hue test. Doing so ensured that these subjects definitely had severe problems of color discrimination in the red-green color spectrum. This procedure is sufficient, since we only wanted to assess the effect of missing color discrimination in the R-G color channel. In principle, we could have also studied protanopes subjects. But due to the much higher incidence of deuteranopia we did not reach any protanopes with our call for participation.

Viewing natural scenes, deuteranope and protanope subjects show no impairments in visual memory compared to color-normal subjects (Gegenfurtner, Wichmann, and Sharpe, 1998).

The authors presented naturally colored and grayscale images of four different categories with presentation times ranging from 50 to 1000ms. There are two possible explanations for the missing difference. Either deuteranope subjects are able to compensate for the reduced chromatic discrimination. Or the chromatic information in the natural scenes examined (Landscapes, Rocks, Flowers and Man-Made objects) is sufficiently represented by the remaining color subsystem. The results of a recent study in our lab (Frey, Honey, & König, 2008) point towards the second explanation in the case of most image categories. We examined the salience of color features in 3 out of these 4 categories. Color-normal subjects were in their selection of fixation points not influenced by R-G color information in Landscapes and Flowers. This could indicate that the fixated or interesting elements of these images are not represented within the R-G color subsystem. Therefore it would be interesting to do the same memory experiment with Rainforest images. We expect a difference between color-normal and dichromatic subjects for presentation times of up to 500 or 600ms. Only after this time we see the maximum influence of R-G color-contrast on overt attention in dichromats.

### **3.5.5 Rainforest images**

The cones of trichromatic primates are optimally tuned to detect food sources in a rainforest environment. These food sources mostly include ripe fruit (Mollon, 1989; Regan, Julliot, Simmen, Vienot, Charles-Dominique, and Mollon, 1998; Sumner and Mollon, 2000) and edible young leaves (Dominy and Lucas, 2001), which have to be detected on a nearly equiluminant background of foliage. This advantage of trichromacy is also reflected in evolutionary selection. While in Caucasians the incidence of R-G color blindness is around 8%, it is less than 0.1% in macaques (Onishi, Koike, Ida, Imai, Shichida, Takenaka, 1999). Therefore it is only too plausible that Rainforest images are one of the classes of stimuli, in which R-G color information has the most influence on overt attention. And indeed, we find R-G contrast to be much higher than B-Y contrast effect sizes. In addition, removing R-G color information significantly alters fixation patterns of color-normal subjects. Taken together, these results show that R-G color information has a considerable influence on attentional processing in this specific environment. However, deuteranope subjects seem able to compensate for their reduced perception of R-G color information by means of top-down strategies.

In summary, our results provide evidence for a causal influence of color contrast on overt attention, which is after a short delay complemented by high level information. It interacts with the image statistics in a way, that in Rainforest with the major variance along the R-G

axis, the feature correlation is highest for this contrast. Nevertheless, evaluation of this color information is done in a hue-invariant fashion. In deuteranopes, mechanisms have developed that compensate for the absence of color information at the cost of a slight delay. As a consequence, in the absence of color information they perform more consistent than normal subjects.

### **3.6 References for Experiments 1-3**

- Açık, A., Onat, S., Schumann, F., Einhäuser, W., & König, P. (submitted). Effects of luminance contrast and its modifications on fixation behavior during free viewing of images from different categories.
- Bacon, W. F., & Egeth, H. E. (1994). Overriding stimulus-driven attentional capture. *Percept Psychophys*, *55*(5), 485-496.
- 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.
- Buswell, G. T. (1935). *How people look at pictures; a study of the psychology of perception in art*. Chicago, Ill: University of Chicago Press.
- Cheal and Lyon (1991). Central and peripheral precuing of forced-choice discrimination. *Q J Exp Psychol A*, *43*(4), 859-880.
- Dayan, P., & Abbott, L. F. (2001). *Theoretical neuroscience: computational and mathematical modeling of neural systems*. Cambridge, Mass.: Massachusetts Institute of Technology Press.
- Derrington, A. M., Krauskopf, J., & Lennie, P. (1984). Chromatic mechanisms in lateral geniculate nucleus of macaque. *J Physiol*, *357*, 241-265.
- Dominy, N. J., & Lucas, P. W. (2001). Ecological importance of trichromatic vision to primates. *Nature*, *410*, 363-366.
- D'Zmura, M. (1991). Color in visual search. *Vision Res*, *31*, 951-966.
- Einhäuser, W., & König, P. (2003). Does luminance-contrast contribute to a saliency map for overt visual attention? *Eur J Neurosci*, *17*, 1089-1097.
- Folk, C. L., Remington, R. W., & Johnston, J. C. (1992). Involuntary covert orienting is contingent on attentional control settings. *J Exp Psychol Hum Percept Perform*, *18*, 1030-1044.
- Frey, H.-P., Wirz, K., Willenbockel, V., Troscianko, T., & König, P. (2006). Do we need the red-green color channel in order to detect fruit? *Society for Neuroscience meeting abstract*.
- Frey, H. P., König, P., & Einhäuser, W. (2007). The role of first- and second-order stimulus features for human overt attention. *Percept Psychophys*, *69*, 153-161.

- Frey, H.-P., Honey, C., & König, P. (2008). What's color got to do with it? The influence of color on overt visual attention in different categories. *J Vis*, 8(14):6, 1–17.
- Gegenfurtner, K. R., Wichmann, F. A., & Sharpe, L. T. (1998). The contribution of color to visual memory in X-chromosome-linked dichromats. *Vision Res*, 38, 1041-1045.
- Gegenfurtner, K. R., & Kiper, D. C. (2003). Color vision. *Annu Rev Neurosci*, 26, 181-206.
- Gottlieb, J. P., Kusunoki, M., & Goldberg, M. E. (1998). The representation of visual salience in monkey parietal cortex. *Nature*, 391, 481-484.
- Helmholtz, H. von (1867). *Handbuch der physiologischen Optik*. Hamburg: Verlag Leopold Voss.
- Hendry, S. H. C., & Reid R. C. (2000). The koniocellular pathway in primate vision. *Annu Rev Neurosci*, 23, 127-153.
- Hoffman, J. E., & Subramaniam, B. (1995). The role of visual attention in saccadic eye movements. *Percept Psychophys*, 57, 787-795.
- Horwitz, G. D., & Newsome, W. T. (1999). Separate signals for target selection and movement specification in the superior colliculus. *Science*, 284, 1158-1161.
- Ignashchenkova, A., Dicke, P. W., Haarmeier, T., & Thier, P. (2004). Neuron-specific contribution of the superior colliculus to overt and covert shifts of attention. *Nat Neurosci*, 7, 56-64.
- Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nat Rev Neurosci*, 2, 194-203.
- Itti, L., Koch, C., & Niebur, E. (1998). A Model of Saliency-Based Visual Attention for Rapid Scene Analysis. *IEEE Trans Pattern Anal Mach Intell*, 20(11), 1254-1259.
- Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: towards the underlying neural circuitry. *Hum Neurobiol*, 4, 219-227.
- Komatsu, H., Ideura, Y., Kaji, S., & Yamane, S. (1992). Color selectivity of neurons in the inferior temporal cortex of the awake macaque monkey. *J Neurosci*, 12, 408-424.
- Krieger, G., Rentschler, I., Hauske, G., Schill, K., & Zetsche, C. (2000). Object and scene analysis by saccadic eye-movements: an investigation with higher-order statistics. *Spatial Vision*, 13, 201-214.
- Kullback, S., & Leibler, R. A. (1951). On Information and Sufficiency. *The Annals of Mathematical Statistics*, 22(1), 79-86.
- Kustov, A. A., & Robinson, D. L. (1996). Shared neural control of attentional shifts and eye movements. *Nature*, 384, 74-77.
- 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. *Eur J Neurosci*, 13, 364-372.

- 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. *Spat Vis*, *10*, 165-188.
- 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.
- Mollon, J. D. (1989). "Tho' she kneel'd in that place where they grew..." The uses and origins of primate colour vision. *J Exp Biol*. *146*, 21-38.
- Müller, H. J., & Rabbitt P. M. A. (1989). Reflexive and Voluntary Orienting of Visual Attention: Time Course of Activation and Resistance to Interruption *J Exp Psychol Hum Percept Perform*, *15*, 315-330.
- Nagy & Sanchez (1990). Critical color differences determined with a visual search task. *J Opt Soc Am A*, *7*, 1209-1217.
- Nathans, J. (1999). The evolution and physiology of human color vision: insights from molecular genetic studies of visual pigments. *Neuron*, *24*, 299-312.
- Onishi, A., Koike, S., Ida, M., Imai, H., Shichida, Y., Takenaka, O., Hanazawa, A., Komatsu, H., Mikami, A., Goto, S., Suryobroto, B., Kitahara, K., & Yamamori, T. (1999). Dichromatism in macaque monkeys. *Nature*, *402*, 139-140.
- Parkhurst, D., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision Res*, *42*, 107-123.
- Parkhurst, D. J., & Niebur, E. (2004). Texture contrast attracts overt visual attention in natural scenes. *Eur J Neurosci*, *19*, 783-789.
- Peters, R. J., Iyer, A., Itti, L., & Koch, C. (2005). Components of bottom-up gaze allocation in natural images. *Vision Res*, *45*, 2397-2416.
- Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annu Rev Neurosci*, *13*, 25-42.
- Regan, B. C., Reffin, J. P., & Mollon, J. D. (1994). Luminance noise and the rapid determination of discrimination ellipses in colour deficiency. *Vision Res*, *34*, 1279-1299.
- Regan, B. C., Julliot, C., Simmen, B., Vienot, F., Charles-Dominique, P., & Mollon, J. D. (2001). Fruits, foliage, and the evolution of primate colour vision. *Philos Trans R Soc Lond B Biol Sci*, *356*, 229-283.
- Reinagel, P., & Zador, A. M. (1999). Natural scene statistics at the centre of gaze. *Network*, *10*, 341-350.
- Rüttiger, L., Braun, D. I., Gegenfurtner, K. R., Petersen, D., Schönle, P., & Sharpe, L. T. (1999). Selective color constancy deficits after circumscribed unilateral brain lesions. *J Neurosci*,

- 19(8), 3094-3106.
- Saal, H., Nortmann, N., Krüger, N., & König, P. (2006). Salient image regions as a guide for useful visual features. *IEEE Advances in Cybernetic Systems*.
- Shevell, S.K., & He, J.C. (1997). The visual photopigments of simple deuteranomalous trichromats inferred from color matching. *Vision Res*, 37, 1115–1127.
- Sumner, P., & Mollon, J. D. (2000). Catarrhine photopigments are optimized for detecting targets against a foliage background. *J Exp Biol*, 203, 1963-1986.
- Tatler, B. W., Baddeley, R. J., & Gilchrist, I. D. (2005). Visual correlates of fixation selection: effects of scale and time. *Vision Res*, 45, 643-659.
- Theeuwes, J. (1994). Stimulus-driven capture and attentional set: selective search for color and visual abrupt onsets. *J Exp Psychol Hum Percept Perform*, 20, 799-806.
- Thompson, K. G., Bichot, N. P., & Schall, J. D. (1997). Dissociation of visual discrimination from saccade programming in macaque frontal eye field. *J Neurophysiol*, 77, 1046-1050.
- Treisman & Gelade (1980). A feature integration theory of attention. *Cognit Psychol*, 12, 97-136.
- 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(ECVP Abstract Supplement).
- Van Zoest, W., Donk, M., & Theeuwes, J. (2004). The role of stimulus-driven and goal-driven control in saccadic visual selection. *J Exp Psychol Hum Percept Perform*, 30, 746-759.
- Yantis, S., & Jonides, J. (1990). Abrupt visual onsets and selective attention: voluntary versus automatic allocation. *J Exp Psychol Hum Percept Perform*, 16, 121-134.
- Zeki, S. (1983). The distribution of wavelength and orientation selective cells in different areas of monkey visual cortex. *Proc R Soc Lond B Biol Sci*, 217, 449-470.
- Zetsche, C. and Barth, E. (1990a). Fundamental limits of linear filters in the visual processing of two-dimensional signals, *Vision Res*, 30, 1111– 1117.

## 4. General Discussion

In order to study the influence of neurophysiologically plausible color features on overt attention, I started with a comparison of different image categories. Examining the salience of various color features I find two extreme categories: while in Rainforest images all color features are salient, none is salient in Fractals. The latter finding contradicts earlier studies that only used a coarse color metric (Parkhurst & Niebur, 2002). No neurophysiologically plausible color feature is salient in all categories of natural scenes. This shows that the influence of color features on overt attention depends on the type of image.

Since Rainforest was the only category in which I found a clear influence of several color features on overt attention, I designed the second series of experiments to examine whether the influence in this category is causal. And indeed, in Rainforest images color-contrast, computed independent of cardinal color axes, causally attracts overt attention in color-normal observers. The computation of such a feature most likely takes place in a higher ventral visual stream area like V4 or in the temporal gyrus ((Zeki, 1983, Rüttiger, Brain, Gegenfurtner, Petersen, Schönle, & Sharpe, 1999). It could, however, also be possible that already in V1 the adequate signals are processed (Li, 2002).

In deuteranope subjects there is a clear influence of higher-level signals on overt attention. One possible explanation supported by my data is that due to the missing color information deuteranope subjects are guided by the objects present in the visual scene. The detection and identification of objects requires higher brain areas like PFC (for a review see: Fenske, Aminoff, Gronau, & Bar, 2006). The configuration of objects in a scene can be used as a top-down cue for control of overt attention.

My thesis provides evidence that different levels of the visual hierarchy are involved in the computation of saliency. This is in line with previous neurophysiological findings in monkeys (Mazer & Gallant, 2003; Bichot, Rossi, & Desimone, 2005). The experiments conducted also highlight the importance of higher-level mechanisms. Even though I could only show direct evidence in deuteranope observers, these mechanisms should also have considerable influence on overt attention in color-normal observers.

## 5. General References

- Averbach, E., & Coriell, A. S. (1961). Short-term memory in vision. *Bell Systems Technical Journal*, *40*, 309-328.
- Bacon, W. F., & Egeth, H. E. (1994). Overriding stimulus-driven attentional capture. *Percept Psychophys*, *55*(5), 485-496.
- 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.
- Blaser, E., Pylyshyn, Z. W., & Holcombe, A. O. (2000). Tracking an object through feature space. *Nature*, *409*, 196-199.
- Buswell, G. T. (1935). *How people look at pictures; a study of the psychology of perception in art*. Chicago, Ill: University of Chicago Press.
- Carmi, R., & Itti, L. (2006). Visual causes versus correlates of attentional selection in dynamic scenes. *Vision Res*, *46*, 4333-4345.
- Derrington, A. M., Krauskopf, J., & Lennie, P. (1984). Chromatic mechanisms in lateral geniculate nucleus of macaque. *J Physiol*, *357*, 241-265.
- Duncan, J. (1984). Selective attention and the organization of visual information. *J Exp Psych Gen*, *113*(4), 501-517.
- D'Zmura, M. (1991). Color in visual search. *Vision Res*, *31*, 951-966.
- Einhäuser, W., & König, P. (2003). Does luminance-contrast contribute to a saliency map for overt visual attention? *Eur J Neurosci*, *17*, 1089-1097.
- Eriksen, C. W., & Hoffman, J. E. (1973). The extent of processing of noise elements during selective coding from visual displays. *Percept Psychophys*, *14*, 155-160.
- Fenske, M. J., Aminoff, E., Gronau, N., & Bar, M. (2006). Top-down facilitation of visual object recognition: object-based and context-based contributions. *Prog Brain Res*, *155*, 3-21.
- Folk, C. L., Remington, R. W., & Johnston, J. C. (1992). Involuntary covert orienting is contingent on attentional control settings. *J Exp Psychol Hum Percept Perform*, *18*, 1030-1044.
- Gegenfurtner, K. R., & Kiper, D. C. (2003). Color vision. *Annu Rev Neurosci*, *26*, 181-206.
- Helmholtz, H. von (1867). *Handbuch der physiologischen Optik*. Hamburg: Verlag Leopold Voss.
- Hendry, S. H. C., & Reid R. C. (2000). The koniocellular pathway in primate vision. *Annu Rev Neurosci*, *23*, 127-153.
- Hillyard, S. A., Vogel, E. K., & Luck, S. J. (1998). Sensory gain control (amplification) as a mechanism of selective attention: electrophysiological and neuroimaging evidence. *Phil. Trans. R. Soc. Lond. B*, *353*, 1257-1270.

- Hoffman, J. E., & Subramaniam, B. (1995). The role of visual attention in saccadic eye movements. *Percept Psychophys*, *57*, 787-795.
- Ignashchenkova, A., Dicke, P. W., Haarmeier, T., & Thier, P. (2004). Neuron-specific contribution of the superior colliculus to overt and covert shifts of attention. *Nat Neurosci*, *7*, 56-64.
- Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nat Rev Neurosci*, *2*, 194-203.
- Kelly, S. P., Gomez-Ramirez, M., & Foxe J. J. (2008). Spatial attention modulates initial afferent activity in human primary visual cortex. *Cereb Cortex*, *18*(11), 2629-2636.
- 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. *Hum Neurobiol*, *4*, 219-227.
- Komatsu, H., Ideura, Y., Kaji, S., & Yamane, S. (1992). Color selectivity of neurons in the inferior temporal cortex of the awake macaque monkey. *J Neurosci*, *12*, 408-424.
- Kustov, A. A., & Robinson, D. L. (1996). Shared neural control of attentional shifts and eye movements. *Nature*, *384*, 74-77.
- Le Meur, O., Le Callet, P., & Barba, D. (2007). Predicting visual fixations on video based on low-level visual features. *Vision Res*, *47*, 2483-2498.
- 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. *Eur J Neurosci*, *13*, 364-372.
- 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. *Spat Vis*, *10*, 165-188.
- Martinez, A., Ramanathan, D. S., John J. Foxe, J. J., Javitt D. C., & Hillyard, S. A. (2007). The role of spatial attention in the selection of real and illusory objects. *J Neurosci*, *27*(30), 7963-7973.
- 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.
- Nagy & Sanchez (1990). Critical color differences determined with a visual search task. *J Opt Soc Am A*, *7*, 1209-1217.
- O'Craven K. M., Downing, P. E., & Kanwisher, N. (1999). fMRI evidence for objects as the units of

- attentional selection. *Nature*, *401*, 584-587.
- Parkhurst, D., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision Res*, *42*, 107-123.
- Peters, R. J., Iyer, A., Itti, L., & Koch, C. (2005). Components of bottom-up gaze allocation in natural images. *Vision Res*, *45*, 2397-2416.
- Posner, M. I., Snyder, C. R. R., Davidson, B. J. (1980) Attention and the detection of signals. *J Exp Psych Gen*, *109*, 160-174.
- Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annu Rev Neurosci*, *13*, 25-42.
- Rizzolatti, G., Riggio, L., Dascola, I., & Umiltà, C. (1987). Reorienting attention across the horizontal and vertical meridians: evidence in favor of a premotor theory of attention. *Neuropsychologia*, *25*, 31-40.
- Roelfsema, P. R., Lamme, V. A. F., & Spekreijse, H. (1998). Object-based attention in the primary visual cortex of the macaque monkey. *Nature*, *395*, 376-381.
- Rüttiger, L., Braun, D. I., Gegenfurtner, K. R., Petersen, D., Schönle, P., & Sharpe, L. T. (1999). Selective color constancy deficits after circumscribed unilateral brain lesions. *J Neurosci*, *19*(8), 3094-3106.
- Saarinen, J. (1993). Focal visual attention and pattern discrimination. *Perception*, *22*(5), 509-515.
- Saenz, M., Buracas, G. T., & Boynton, G. M. (2002). Global effects of feature-based attention in human visual cortex. *Nat Neurosci*, *5*(7), 631-632.
- Saenz, M., Buracas, G. T., & Boynton, G. M. (2003). Global feature-based attention for motion and color. *Vision Res*, *43*, 629-637.
- Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs: General and Applied*, *74*(11), 1-30.
- Tatler, B. W., Baddeley, R. J., & Gilchrist, I. D. (2005). Visual correlates of fixation selection: effects of scale and time. *Vision Res*, *45*, 643-659.
- Theeuwes, J. (1994). Stimulus-driven capture and attentional set: selective search for color and visual abrupt onsets. *J Exp Psychol Hum Percept Perform*, *20*, 799-806.
- Treisman & Gelade (1980). A feature integration theory of attention. *Cognit Psychol*, *12*, 97-136.
- Valdes-Sosa, M., Bobes, M. A., Rodriguez, V., & Pinilla, T. (1998). Switching attention without shifting the spotlight: object-based attentional modulation of brain potentials. *J Cog Neurosci*, *10*(1), 137-151.
- Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided Search: an alternative to the feature integration model for visual search. *J Exp Psychol Hum Percept Perform*, *15*, 419-433.

Yarbus, A. L. (1967). *Eye movements and vision*. New York: Plenum Press.

Yeshurun, Y., & Carrasco, M. (1998). Attention improves or impairs visual performance by enhancing spatial resolution. *Nature*, *396*, 72-75.

Zeki, S. (1983). The distribution of wavelength and orientation selective cells in different areas of monkey visual cortex. *Proc R Soc Lond B Biol Sci*, *217*, 449-470.

## 6. Acknowledgements

First and foremost, I would like to thank my supervisor Peter König. Without him, this whole thesis wouldn't have been possible. His enthusiasm, support, and critical thinking were invaluable.

I would like to thank Tom Troscianko for providing the Rainforest images and a lot of good advice, as well as Wolfgang Einhäuser for great discussions.

Furthermore, I would like to thank all the co-authors and students that have contributed to the publications: Kerstin Wirz, Verena Willenbockel, Sepp Kollmorgen, Sylvia Schröder, Christian Honey, Cornell Schreiber, and Torsten Betz.

A big thank-you goes to my parents for their constant love and support.

Especially, I would like to thank my wife Anita. She did so much for me during those years.