

Recognising the forest, but not the trees: An effect of colour on scene perception and recognition

Tanja C.W. Nijboer ^{a,*}, Ryota Kanai ^b, Edward H.F. de Haan ^a,
Maarten J. van der Smagt ^a

^a *Utrecht University, Helmholtz Institute, Department of Experimental Psychology, Heidelberglaan 2, 3584 CS Utrecht, The Netherlands*

^b *California Institute of Technology, Division of Biology, MIC 114-96, Pasadena, CA 91125, USA*

Received 23 March 2007

Available online 22 August 2007

Abstract

Colour has been shown to facilitate the recognition of scene images, but only when these images contain natural scenes, for which colour is ‘diagnostic’. Here we investigate whether colour can also facilitate memory for scene images, and whether this would hold for natural scenes in particular. In the first experiment participants first studied a set of colour and greyscale natural and man-made scene images. Next, the same images were presented, randomly mixed with a different set. Participants were asked to indicate whether they had seen the images during the study phase. Surprisingly, performance was better for greyscale than for coloured images, and this difference is due to the higher false alarm rate for both natural and man-made coloured scenes. We hypothesized that this increase in false alarm rate was due to a shift from scrutinizing details of the image to recognition of the gist of the (coloured) image. A second experiment, utilizing images without a nameable gist, confirmed this hypothesis as participants now performed equally on greyscale and coloured images. In the final experiment we specifically targeted the more detail-based perception and recognition for greyscale images versus the more gist-based perception and recognition for coloured images with a change detection paradigm. The results show that changes to images are detected faster when image-pairs were presented in greyscale than in colour. This counterintuitive result held for both natural and man-made scenes (but not for scenes without nameable gist) and thus corroborates the shift from more detailed processing of images in greyscale to more gist-based processing of coloured images.

© 2007 Elsevier Inc. All rights reserved.

Keywords: Colour; Memory; Gist; Details; Colour diagnosticity; Change detection

1. Introduction

Humans are generally very good at visual recognition. A number of studies have documented our ability to identify hundreds of familiar objects and faces in a crowd of thousands of unfamiliar counterparts. In

* Corresponding author.

E-mail address: t.c.w.nijboer@fss.uu.nl (T.C.W. Nijboer).

addition, Shepard (1967) and Standing (1973) have demonstrated that we have an enormous capacity for remembering and recognising pictures. From a more ecological point of view, it is surprising that memory for scene images has not attracted more scientific scrutiny. A scene is usually defined as a semantically coherent, nameable view of an environment, composed of multiple discrete objects. The studies that have looked at scene recognition often report that human visual memory may be limited in capacity and detail. For example, replacements of objects within a scene—between encoding and recognition—frequently go undetected (Friedman, 1979).

One explanation for these diverse observations is that visual information may be stored in more than one manner. It has been suggested that the image of a familiar scene can be stored in visual memory while the concept (or ‘gist’) may be encoded simultaneously in semantic memory. The term ‘gist’ generally refers to an interpretation of a story or an image on the basis of existing knowledge (Friedman, 1979; Luck & Vogel, 1997; Treisman, 1986). Depending on the conditions, the gist of an image may be remembered better than the visual details (Friedman, 1979; Miller & Gazzaniga, 1998; Rensink, 2000) or vice versa.

The proficiency of visual memory depends on the time allowed to scrutinize details and the extent to which details are recognisable entities. Gist perception is considered to be fast (Biederman, 1981; Intraub, 1980; Oliva, 2005; Oliva & Schyns, 1997; Potter, 1975; Potter, 1976). For example, Potter (1975), Potter (1976) demonstrated that less than 125 ms is needed for the meaning of an image to be understood. An interaction between visual scrutiny and gist recognition is suggested by the observation that expected or scene-consistent objects have been found to receive less attention, and subsequently, are encoded, stored, and retained in memory in a less detailed fashion than unexpected or inconsistent objects (Friedman, 1979; Hollingworth & Henderson, 2002).

An important visual cue that may affect both visual memory and gist recognition is colour. Only a few studies have specifically investigated the role of colour in memory for scene images. Homa and Viera (1988) found an advantage of coloured over black-and-white images across different retention intervals, ranging from immediate recall to a 12-week delay. They concluded that information about surface characteristics, such as colour, enhances memory. The effect of colour is modulated by the type of scene. In natural scenes (e.g. coasts, deserts, forests), there is a more consistent colour pattern that serves as a recognition cue compared to man-made scenes (e.g. cities, markets, rooms). In man-made objects and scenes, object-colour relationships are essentially random. Natural scenes, with a more consistent colour pattern, are also known as scenes high in ‘colour diagnosticity’ and have been found to be more (negatively) influenced by colour manipulations (e.g. inverted colour, greyscale) than scenes low in colour diagnosticity in recognition and naming studies (Nijboer, van der Smagt, van Zandvoort, & de Haan, 2007; Oliva & Schyns, 1997; Tanaka & Presnell, 1999). It has been suggested that the effect of colour takes place at the level at which knowledge of object properties is stored (Humphrey, Goodale, Jakobson, & Servos, 1994; Nijboer et al., 2007). In other words, colour helps us to recognise natural scenes, because it facilitates the identification of the gist of colour diagnostic images.

More recently, Wichmann, Sharpe, and Gegenfurtner (2002) investigated the role of colour in recognition memory with carefully luminance matched colour and greyscale images. They used four categories of images: green landscapes, flowers, rock formations, and man-made objects. They found an overall 5–10% enhanced performance for the colour images compared to the greyscale images, irrespective of presentation duration. They concluded that colour as a surface property, at least for natural scenes, is stored in memory, but that sensory facilitation also played an important role, as colour recognition superiority was found independent of colour diagnosticity. Natural and man-made scenes, however, were not equally distributed in the study of Wichmann et al. (2002), with fewer images of man-made scenes. The present study, therefore, was aimed specifically at the role of colour in recognising natural and man-made scenes. A delayed-matching-to-sample method will be used to investigate whether colour information improves scene memory for natural (colour diagnostic) scenes more than man-made scenes. In order to be able to compare our results directly, we designed a comparable experiment to the one used by Wichmann et al. (2002).

2. Experiment 1

The aim of this experiment was to investigate the effect colour and the recognition of scenes. We used a delayed-matching-to-sample design with coloured and greyscale images of natural and man-made scenes. In

order to make a direct comparison with [Wichmann et al. study \(2002\)](#), we also included a condition with inverted colour and presented images for a set of different durations.

2.1. Method

2.1.1. Participants

Twenty students (mean age 25.90, SD 3.05) participated in this study. All had normal or corrected-to-normal vision and reported no colour blindness, which was confirmed by the [Ishihara test for colour blindness \(1977\)](#).

2.2. Apparatus, stimuli, and procedure

The experiment was programmed using Matlab and the Psychophysics Toolbox ([Brainard, 1997; Pelli, 1997](#)). Stimuli were presented on a 21 in. monitor (80 Hz, 1024 × 768). Viewing distance was 57 cm which was controlled for with a chin and forehead rest.

Images were chosen from different categories: coasts, forests and deserts for the natural scenes, and cities, markets and rooms for the man-made scenes. As [Wichmann et al. \(2002\)](#) found that recognition performance (HR and d') depends on image category, we used a comparable set of different scenes within *each* scene type (three different natural scenes and three different man-made scenes).

All images were presented in colour, inverted colour and greyscale. Inverted colour images as well as greyscale images were manipulated in Adobe Photoshop from the original colour images. Images were first screened for category membership by 25 students, who did not participate in any of the experiments. Only images that were categorised consistently by all students were included. The mean luminance was identical for all image formats. The size of the images was 256 × 256 pixels, subtending approximately 8° × 8° of visual angle. [Fig. 1](#) gives some examples of natural and man-made scenes in the different image formats used.

The participants were tested individually in a quiet, darkened room. The experiment consisted of two phases: first, a study phase, in which the participants were sequentially shown a set of 96 images (32 in colour, 32 in inverted colour, 32 in greyscale), one at the time in a random order; second, a test phase, in which the participants' recognition memory for images was tested. In the study phase, participants only had to watch the images. Presentation durations of the images were varied from 50 ms up to 800 ms (50, 200, 400, 800 ms) and images were randomly assigned to one of the four presentation durations. In between two successive images, a



Fig. 1. Some examples of the images used in Experiment 1.

blank screen was presented for 50 ms. This inter-stimulus interval (ISI) was chosen, as Potter (1976) showed that recognition memory for scenes was strongly affected by SOA in rapid serial visual presentation (RSVP) of scenes. With our relatively short ISIs, SOAs are now mainly determined by the stimulus duration, hence we expected a strong effect of stimulus duration on recognition memory for scenes.

In the test phase, participants had to indicate whether the image shown was presented in the study phase and respond by pressing a key on the keyboard. The test phase started 3 min after the study phase had ended. The same 96 images were randomly mixed with 96 new images of the same scene categories.

2.3. Analysis

Repeated measures ANOVAs were performed on hit rate (HR), false alarm rate (FAR), discriminability (d'), criterion (c), and bias (β).¹ Since no interactions between presentation durations and any measures were found (p 's for all measures $>.100$),² the data were collapsed across duration. Hit rate, false alarm rate, d' , criterion, and bias were calculated per scene type (natural versus man-made), per format (colour, inverted colour, greyscale) for each participant individually. For post hoc testing, pair-wise comparisons with Bonferroni correction were used.

2.4. Results and discussion

Table 1 shows the mean discriminability, criterion, and response bias for natural and man-made scenes, split by image format. As can be seen from Table 1, d' was influenced by the format of the presented images ($F(2, 19) = 26.320, p < .001$), with the highest discriminability for the greyscale images for both the natural ($p < .001$; for greyscale images versus both colour and inverted colour images) and the man-made scenes ($p < .001$; for greyscale images versus both colour and inverted colour images). We have also analysed the decision criterion and response bias (see Table 1). Decision criterion and bias shifted depending on the image format ($F(2, 19) = 40.609, p < .001$), and ($F(2, 19) = 46.336; p < .001$), respectively. These shifts in decision criteria and bias indicate that participants responded more conservatively to the greyscale images as compared to colour ($p < .001$) and inverted colour images ($p < .001$).

Fig. 2 shows the mean hit rate and false alarm rate per scene type (natural and man-made), split by image format (colour, inverted colour, and greyscale). As can be seen from Fig. 2, the influence of format on hit rate is different for natural and man-made scenes ($F(2, 19) = 4.389, p < .020$). With natural scenes, colour images have a significantly higher hit rate than both inverted colour images ($p < .001$) and greyscale images ($p < .020$). There is no significant difference between greyscale and inverted colour images. With man-made scenes, hit rate is not influenced by format. The influence of format on false alarm rate was the same for natural and man-made scenes. With both natural and man-made scenes, the false alarm rate for greyscale images is much lower than for both colour and inverted colour images ($p < .001$). In addition, with natural scenes, the false alarm rate for colour and inverted colour images differ significantly ($p < .001$).

Taken together, recognition memory in terms of d' is best for greyscale images. This difference in performance is largely due to a significantly lower false alarm rate for the greyscale images than for the colour images. There was no difference in hit rate between coloured and greyscale man-made scenes, whereas hit rate was larger for coloured than greyscale images with natural scenes. These findings contradict the results reported by Wichmann et al. (2002) in two ways. First, they observed a better overall performance with coloured images and second, they did not find a difference in false alarm rate between coloured natural and man-made images. We suggest that the better performance with greyscale images was due to the fact that coloured images lead to faster identification of the gist of the scene, which in turn might have resulted in a less thorough scrutiny of the visual details of the to be remembered image. The reason this effect did not happen in the Wichmann et al. (2002) experiment might result from the longer intervals between images during the study phase. In

¹ $d' = |z \text{ hit rate} - z \text{ false alarm rate}|$; $c = -0.5 * |z \text{ hit rate} + z \text{ false alarm rate}|$; $\beta = 0.5 * d' * |z \text{ hit rate} + z \text{ false alarm rate}|$.

² Although there were no significant interaction, we did find a significant main effect of presentation duration; d' , c , and β for images shown for only 50 ms differed from all other presentation durations (200, 400, 800 ms; $p < .001$), and d' for images shown for 200 ms differed from the images shown for 800 ms ($p < .010$).

Table 1
 d' , criterion, and bias per image format for recognition of different scene categories and image formats

	Natural colour	Inverted colour	Greyscale	Man-made colour	Inverted colour	Greyscale
d'	1.020	0.767	1.547	1.052	0.961	1.844
Criterion	-0.654	-0.690	-1.396	-0.809	-0.735	-1.613
Bias	0.878	1.280	2.318	1.314	0.945	2.697

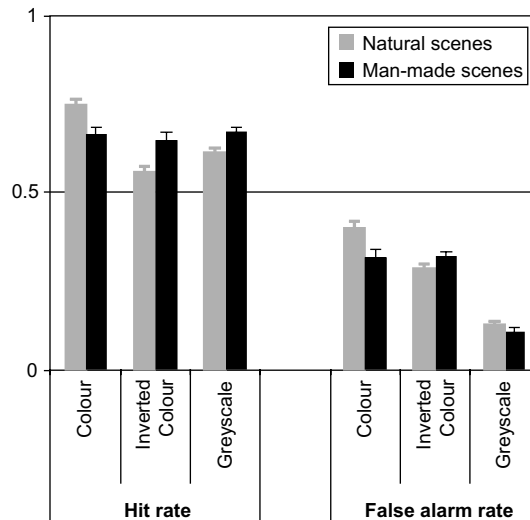


Fig. 2. Effect (hit rate and false alarm rate; error bars depict ± 1 SEM) of image format on recognition of different scene types.

other words, high false alarm rates for colour images might be obtained only when the test phase requires participants to distinguish between visually similar targets and distractors. This explanation is reminiscent of the experimental work on ‘false memories’. This concerns the effect of *expectation* on memory for gist. Participants are more likely to ‘remember’ objects that were not present in the original stimulus when they are semantically associated with the gist of the image (Brewer & Treyans, 1981; Loftus, 2003). Potter, Staub, and O’Connor (2004) demonstrated that false recognition of an image occurred more frequently when images with a conceptually similar gist were shown during the study phase, suggesting that stored semantic knowledge can interfere with detailed visual memory.

A test of our hypothesis that it is the fast identification of the gist that is responsible for the higher false alarm rate in coloured images, is the use of images that do not contain a readily nameable gist. For this second experiment, we used close-up images of tree bark or pebbles instead of pictures of scenes.

3. Experiment 2

This experiment investigates the effect of coloured and greyscale images of different textures on recognition memory. We did not use inverted coloured images as there are no a-priori correct colour-texture relationships for these stimuli.

3.1. Methods

3.1.1. Participants

Sixteen students (mean age 27.20, SD 3.85) participated in this study. All had normal or corrected-to-normal vision and reported no colour blindness, which was confirmed by the Ishihara test for colour blindness (1977).

3.2. Apparatus, procedure, and stimuli

Apparatus and procedure were identical to Experiment 1. Images that do not contain a specific nameable gist were selected for this experiment. We chose close-up images of for example tree bark, leaves, and pebbles (see Fig. 3 for some examples). Seventy-two images were presented in colour and greyscale (36 colour, 36 greyscale) in the study phase and were interleaved with 72 distracter images during the test phase. The size of the images was 256×256 pixels, subtending approximately $8^\circ \times 8^\circ$ of visual angle. Viewing distance was 57 cm.

3.3. Analysis

Hit rate (HR) and false alarm rate (FAR) were calculated per format for each participant individually, and were submitted to repeated measures of ANOVA.

3.4. Results and discussion

As can be seen in Fig. 4, there was no difference in hit rate ($F(1, 15) = 2.98, p > .05$) nor false alarm rate ($F(1, 15) = .456, p > .5$) between coloured and greyscale images. These results show that when images are presented without a nameable gist, recognition memory is comparable for coloured and greyscale images. This finding suggests that colour information per se is not sufficient for the increase in false alarm rate observed in Experiment 1 with coloured images.

The observed increase in false recognition for coloured images in Experiment 1 could have two (not mutually exclusive) origins. First, when coloured images are presented in the *test phase*, the relatively fast gist recognition might compel the participant to rely more on gist than on details, when comparing these images to images stored in memory. In such a situation it might be more difficult to discard an image with the same gist

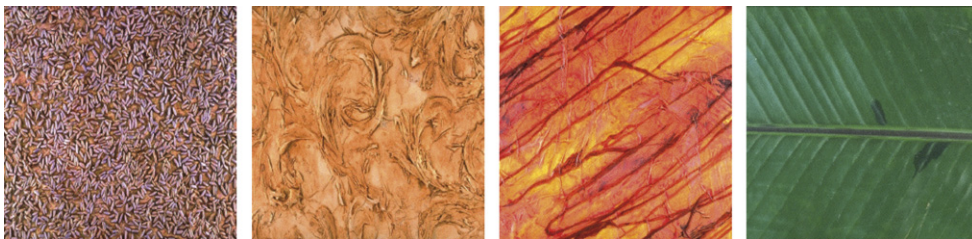


Fig. 3. Some examples of the images used in Experiment 2.

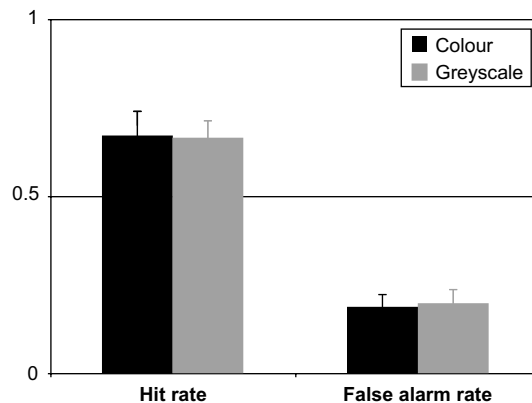


Fig. 4. Effect (hit rate and false alarm rate; error bars depict ± 1 SEM) of image format on recognition of images with no specific nameable content.

as other (different) images that have been presented during the study phase. This account places the effect at the *semantic* stage of memory processing. From the work of Loftus (2003) and Potter et al. (2004), it is likely that semantic confusion plays a role. The extent to which such confusions will occur depends on the semantic content and number of the other images to be remembered, and the number and content of the distracters at the test phase. Second, when a coloured image of a natural scene is presented for later recognition (study phase), the fast recognition of the gist could lead to less detailed encoding of the image. This is an explanation at the *pictorial* level. This account appears less likely but has not been investigated before. Therefore, in order to investigate the possibility that differential visual scrutiny may play a role, we designed a change detection task with both coloured and greyscale images.

4. Experiment 3

In this experiment, we test whether the detection of change, as index of the proficiency of visual analysis, is the critical factor for the difference between coloured and greyscale images. If recognition of greyscale images is indeed based on details rather than gist, change detection performance should be faster for greyscale images than for colour images. This hypothesis is based upon existing explanations of change detection phenomena (Rensink, 2002). It has been shown that obvious changes to images can be hard to detect unless these changing (aspects of) objects are essential to the gist of the image (or denoted as “of central interest” to the image; see Rensink, O’Regan, & Clark, 1997). For example, the disappearance of an engine from the picture of an aeroplane is difficult to detect, because an aeroplane without an engine is still recognised as an aeroplane. In contrast, it is easier to detect the disappearance of a wing, because wings are an essential feature of an aeroplane. In this study, we included coloured and greyscale images of natural and man-made scenes. A separate control condition contained images without a nameable gist.

4.1. Method

4.1.1. Participants

This experiment was set up with a between subject design, since participants in this experiment could only be shown half of the image-pairs (see below). Thirty-two participants (mean age 33.4, SD 2.7) were included in this experiment and 20 additional participants (mean age 30.7, SD 2.8) for the separate control condition. All had normal or corrected-to-normal vision and reported no colour blindness, which was confirmed by the Ishihara test for colour blindness (1977).

4.2. Stimuli and procedure

Scene image-pairs were chosen from different categories: coasts, forests and deserts for the natural scenes, and cities, markets and rooms for the man-made scenes. None of the images were used in the other experiments. The mean luminance was identical for all image formats. The size of the images was 256×256 pixels, subtending approximately $8^\circ \times 8^\circ$ of visual angle. The change between the first and the second image concerned details, such as replacements or deletions of whole objects, or changes in position of objects. Participants were informed before the start of the experiment about the nature of the changes and were specifically instructed that a change never solely consisted of a mere colour change of an object or part of the scene (as this would have been hard to find in the greyscale image-pairs). Changes were equally distributed across the images (left, right, upper part, lower part). Fig. 5 shows examples of the kind of changes that were incorporated.

Each image pair was presented both in colour and in greyscale, but participants never viewed the same scene twice. The experiment entailed 24 scene image-pairs, presented in an alternating fashion with 200 ms presentation duration for each image of the pair, interleaved by a 200 ms blank interval. The sequence was repeated until the participant indicated (by pressing the space bar) that he/she had noticed the difference between the first and the second image of the image pair. Subsequently, they had to verbally describe the change they detected. Image-pairs were shown until response or 2 min at most.

Two sets of image-pairs were used. The first set contained 12 scenes of the 24 total in greyscale and the other 12 in colour. The second set contained the same scenes, but image format was reversed. Half of the



Fig. 5. Some examples of the image-pairs used in Experiment 3. Top: scenes with a readily nameable gist, bottom: scenes without a readily nameable gist.

participants viewed one set; the other participants viewed the other set. All scenes were thus viewed 16 times in both formats.

The control condition also contained 24 scenes (close-up images-pairs of for instance walls, pebbles, or sweets and abstract figures), divided into two sets as outlined above. Again half of the participants viewed one set, while the remaining participants viewed the other. All control scenes were thus viewed 10 times in both formats.

4.3. Analysis

The time from stimulus-onset to the participant's key press (reaction time, RT) was recorded. Across participants, this resulted in RT-distributions for the different image formats (colour and greyscale) and scene types (natural and man-made scenes). Cases where participants did not describe the change correctly or failed to detect the change ($RT > 2$ min) were excluded from the distributions. Since RT-distributions are notoriously skewed, the resulting distributions were subsequently compared using the Kolmogorov–Smirnov (K–S) test. In addition, cumulative distributions were plotted (fraction of changes detected as a function of RT) and psychometric functions (Weibull-variant) were fitted through the data. The 95% confidence intervals for the fit-parameters (notably scale and shape parameters) were calculated and compared.

4.4. Results and discussion

In more than 85% of the trials the change was detected within 30 s presentation duration. Fig. 6 (top) shows the median RT (left panel) for both image formats. The overall finding is that changes in greyscale images were detected faster than in coloured images (K–S $Z = 2.23$, $p < .001$). The right panel shows the cumulative distribution plots (circles are individual trials, the line depicts the fitted psychometric function). Both psychometric functions were excellent fits ($R^2 > 0.99$). The 95% confidence intervals of the scale parameter of both functions did not overlap (greyscale: 4.17–4.94, colour: 7.25–7.71), indicating again a different underlying change detection RT-distribution for greyscale and colour image formats. These results corroborate our hypothesis that in coloured images the encoding emphasis is shifted towards encoding of gist compared to greyscale images, where more details are encoded.

For the control condition (Fig. 6, bottom), there was no significant RT-difference between formats (K–S $Z = 0.74$, $p = .647$). The 95% confidence intervals for the psychometric fit-parameters ($R^2 > 0.98$) of the two image formats showed considerable overlap as well. Even though statistical power in the control condition

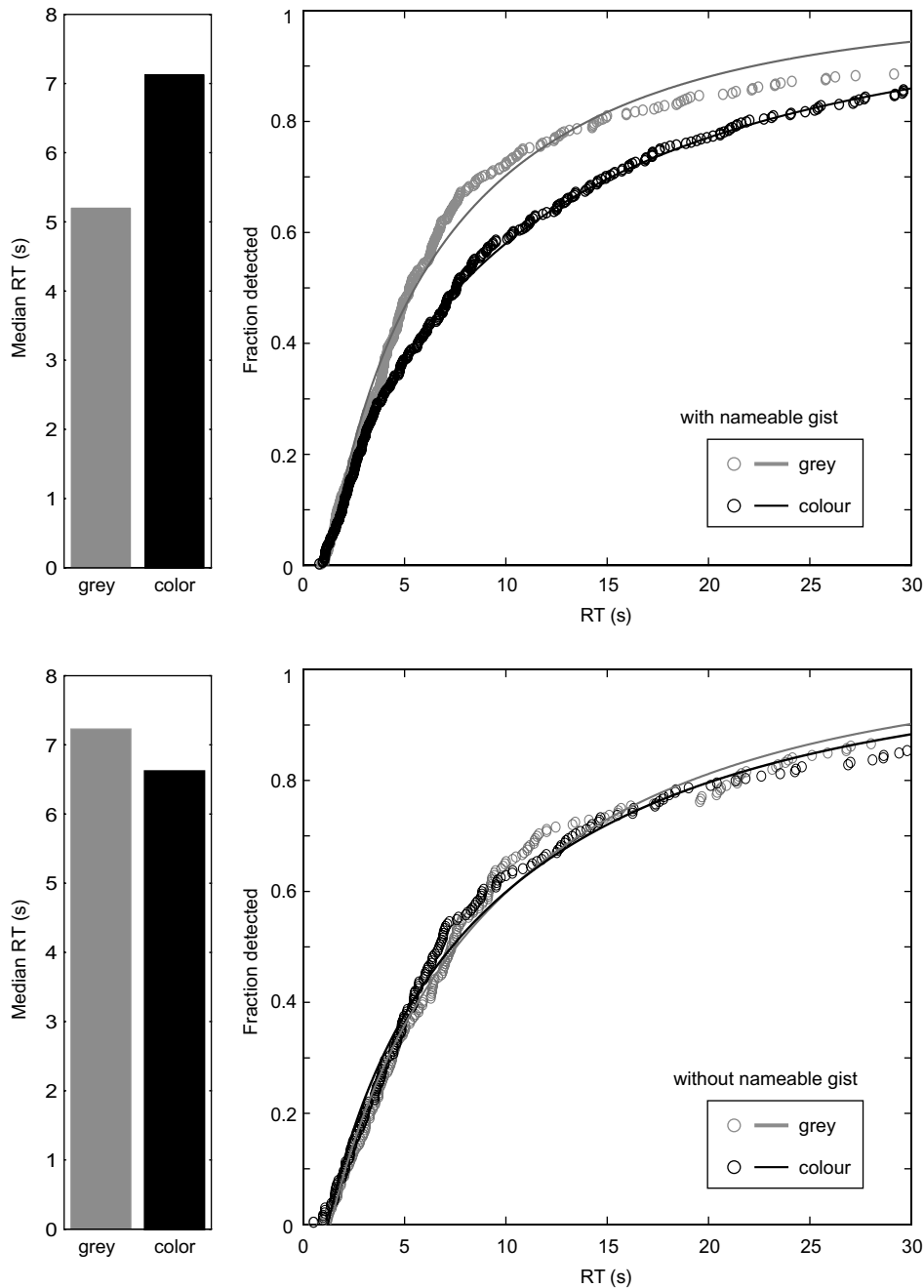


Fig. 6. Median reaction times and cumulative distribution of reaction times, split on image format. Top: scenes with a readily nameable gist, bottom: scenes without a readily nameable gist.

was slightly lower (10 versus 16 presentations per image per format), in the analysis shown below only eight presentations (per image per format) was sufficient to yield significant differences in the main conditions. Moreover, if anything, more power would have resulted in changes to be detected *slower* in greyscale images, compared to coloured images (see for instance Fig. 6 bottom, left panel).

The images used in this experiment included man-made as well as natural scene categories. Thus the RT-distributions were also compared separately for both scene types. The results are shown in Fig. 7. Again, the

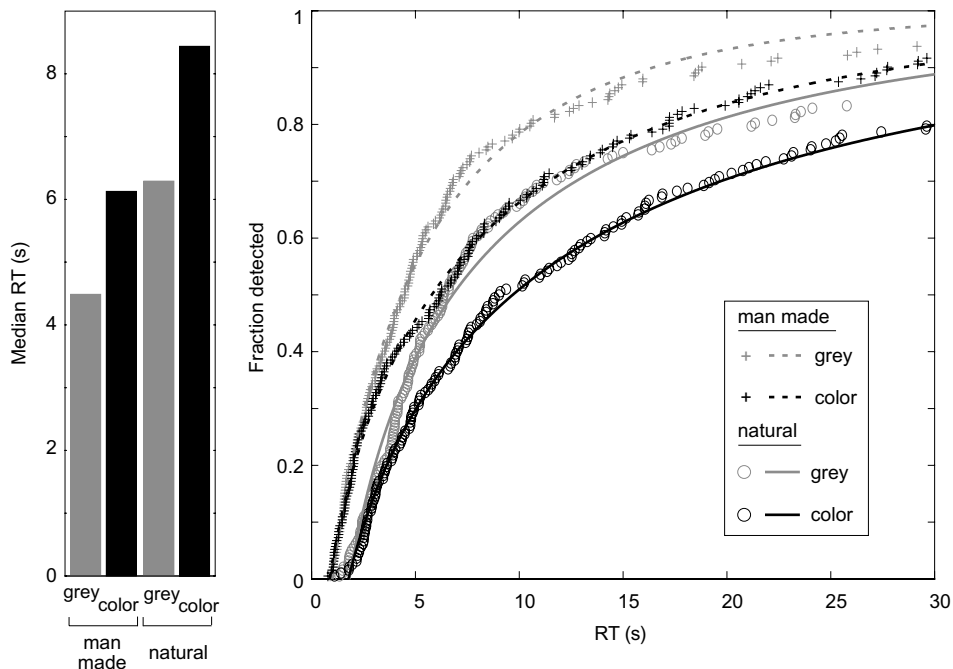


Fig. 7. Median reaction times and cumulative distribution of reaction times, split on scene type and image format.

left panel shows the median RT for greyscale and colour images, for man-made as well as natural scenes. It is immediately clear that for both types of scene the changes in greyscale were detected faster than changes in colour (man-made: $K-S Z = 1.59$, $p = .013$; natural: $K-S Z = 1.64$, $p = .009$). The right panel again shows the cumulative distribution plots with fitted psychometric functions ('+' and dotted lines show data and fits for man-made scenes, circles and continuous lines show the data and fits for natural scenes). All psychometric functions were excellent fits ($R^2 > 0.98$). The 95% confidence intervals of the scale parameter overlapped between scene category, but did not overlap between image format (man-made greyscale: 3.43–4.00, man-made colour: 5.27–5.84; natural greyscale: 1.84–3.67, natural colour 5.70–7.03). In addition the 'shape' parameter differed between formats for the man-made scenes and between scene types (95% CI: man-made greyscale 0.64–0.71, man-made colour 0.57–0.61; natural greyscale 0.41–0.51, natural colour 0.47–0.51).

These analyses show that for both natural and man-made scenes changes in greyscale images are generally detected faster than changes in coloured images, but only when these images contain a discernable and nameable gist. In addition, they seem to indicate that changes in man-made scenes are detected sooner than changes in natural scenes. However, the latter result might easily be attributed to the differences in complexity between images from the two categories. Moreover, in this comparison only 12 man-made and 12 natural scenes were included, which were arbitrarily chosen. These images cannot readily be compared. However, this does not influence our main result (greyscale images versus coloured images), since this comparison deals with *the same* scenes, presented either in greyscale or in colour.

5. General discussion

In the current study, we investigated memory for scene images. Our main research question concerned the interaction between the effect of identification of the scene (gist extraction) and the effect of coloured versus greyscale images (colour diagnosticity). In contrast to previous reports, we found that greyscale images were remembered better than coloured ones. More detailed analysis showed, as expected, that the hit rate was higher for coloured than greyscale images of natural scenes. With man-made scenes, the hit rate was not influenced by format. More surprisingly, however, we found that false alarm rate was substantially higher with

coloured images for both scene types. This means that recognition memory is worse for coloured images because participants are prone to accept distractors more easily.

The discrepancy between our findings and those of [Wichmann et al. \(2002\)](#) might be the result of differences in inter-stimulus interval (ISI): our inter-stimulus interval (ISI) was 50 ms, whereas in the experiment of [Wichmann et al. \(2002\)](#) the ISI was 7000 ms. This is a crucial difference, as participants in the study of [Wichmann et al. \(2002\)](#) had more time between successive images to consolidate more image details. It appears then that consolidation of image details results in fewer false alarms, which will affect performance with the coloured images more than with greyscale images. In other studies, it has been found that scene memory builds up across separate glances and over a period of minutes ([Hollingworth, 2005](#); [Hollingworth & Hollingworth, 2004](#); [Melcher, 2001](#); [Melcher, 2006](#); see also the results of our Experiment 3). Therefore, it might have been more difficult for participants in the present study to remember the details of individual images, which could have resulted in differences in false recognition. Thus, we suggest that our observation that colour hampers recognition memory for scenes arises in experimental conditions in which encoding is restricted by time constraints. Other differences between our study and [Wichmann et al. \(2002\)](#) are the number of categories (six categories (three natural, three man-made) versus four categories (three natural, one man-made), respectively) and the number of images used (96 versus 48, respectively). Each of these differences might have contributed to the differences between our study and [Wichmann et al. \(2002\)](#).

In Experiment 2, we addressed the question whether identification of the scene is a necessary condition for the paradoxical colour effect to appear. To that end, we used images that did not contain a specific nameable content (i.e. gist). Under these conditions there were no differences in memory performance between coloured and greyscale formats (equal hit rate and false alarm rate). This finding strongly suggests that the increase in false alarms as found in Experiment 1 is contingent on the identification of the scene, and not on colour per se.

[Potter et al. \(2002, 2004\)](#) suggested that there are two components of visual memory, a pictorial and a conceptual store, with a faster decay rate for the former compared to the latter. This idea for a multiple component memory system account is further supported by [Cooper and Schacter \(1993\)](#) who suggest that achromatic, detailed structural descriptions and more general, internalised gist-based descriptions in which colour information is also stored can be distinguished. On the basis of our findings in Experiments 1 and 2, we argued that—under time restrictions—colour information modulates the balance to which these two memory components are engaged. In Experiment 3, we tested the counterintuitive hypothesis that colour deteriorates the visual exploration of natural and man-made scenes. The results showed that changes are detected faster in greyscale compared to coloured images, when these changes are unrelated to the gist of the scenes. Our results concur with a recent study by [Torralba, Oliva, Castelhamo, and Henderson \(2006\)](#). They suggest that attention is directed to gist-related locations in a more or less automatic fashion, in which case attention is diverted from the details where the changes occur in our experiment. If colour, as we propose, induces a shift towards a more gist-based encoding of an image, attention will as a consequence be directed even stronger towards these gist-related locations. In the control condition of Experiment 3, we used abstract images without a readily nameable gist and found that there was no difference between the coloured and greyscale images, which strengthens our interpretation that it is the combination of colour and gist and not merely colour, that is responsible for these results. These findings of the change detection experiments are in accordance with the results from Experiments 1 and 2, where the false alarm rate was overall higher for coloured images in both natural and man-made scene types, but not for abstract coloured images without a readily nameable gist.

In sum, we have shown that recognition memory for scenes is influenced by identification or gist extraction. In terms of a two component model of visual memory, entailing a pictorial and a gist-based code, colour modulates the degree to which these two codes are employed. This effect is more pronounced under time restrictions. In our view, coloured images, and especially coloured images of natural scenes, lead to faster gist identification and to less detailed visual scrutiny. The latter effect becomes apparent when participants are required to detect changes as fast as possible. Detection speed is significantly delayed when images are presented in colour.

Acknowledgments

We thank Harald Kunst, Ignace Hooge, and Chris Paffen for the many useful discussions.

References

- Biederman, I. (1981). On the semantics of a glance at a scene. In M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual organization*. Hillsdale, NJ: Erlbaum Associates Inc.
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, 10, 433–436.
- Brewer, W. F., & Treyans, J. C. (1981). The role of schemata in memory for places. *Cognitive Psychology*, 13, 207–230.
- Cooper, L. A., & Schacter, D. L. (1993). Implicit and explicit memory for novel visual objects: Structure and function. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19(5), 995–1009.
- Lofthman, A. (1979). Framing pictures: The role of knowledge in automatized encoding and memory for gist. *Journal of Experimental Psychology: General*, 108(3), 316–355.
- Hollingworth, A., & Henderson, J. M. (2002). Accurate visual memory for previously attended objects in natural scenes. *Journal of Experimental Psychology: Human Perception and Performance*, 28, 113–136.
- Homa, D., & Viera, C. (1988). Long-term memory for pictures under conditions of thematically related foils. *Memory and Cognition*, 16(5), 411–421.
- Hollingworth, A. (2004). Constructing visual representations of natural scenes: The roles of short- and long-term visual memory. *Journal of Experimental Psychology: Human Perception and Performance*, 30(3), 519–537.
- Hollingworth, A. (2005). The relationship between online visual representation of a scene and long-term scene memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(3), 396–411.
- Humphrey, G. K., Goodale, M. A., Jakobson, L. S., & Servos, P. (1994). The role of surface information in object recognition: Studies of a visual form agnostic and normal subjects. *Perception*, 23, 1457–1481.
- Intraub, H. (1980). Presentation rate and the representation of briefly glimpsed pictures in memory. *Journal of Experimental Psychology*, 6(1), 1–12.
- Ishihara, S. (1977). *Tests for colour-blindness* (38 plates ed.). Tokyo, Japan: Kanehara Shuppan Co., Ltd.
- Loftus, E. F. (2003). Make-believe memories. *American Psychologist*, 58, 867–873.
- Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, 390, 279–281.
- Melcher, D. (2001). Persistence of visual memory for scenes. *Nature*, 412, 401.
- Melcher, D. (2006). Accumulation and persistence of memory for natural scenes. *Journal of Vision*, 6(1), 8–17.
- Miller, M. B., & Gazzaniga, M. S. (1998). Creating false memories for visual scenes. *Neuropsychologia*, 36(6), 513–520.
- Nijboer, T. C. W., van der Smagt, M. J., van Zandvoort, M. J. E., & de Haan, E. H. F. (2007). Colour agnosia deteriorates the recognition of natural but not non-natural scenes. *Cognitive Neuropsychology*, 24(2), 152–161.
- Oliva, A. (2005). Gist of the scene. In L. Itti, G. Rees, & J. K. Tsotsos (Eds.), *Neurobiology of attention* (pp. 251–256). San Diego, CA: Elsevier.
- Oliva, A., & Schyns, P. G. (1997). Coarse blobs or fine edges? Evidence that information diagnosticity changes the perception of complex visual stimuli. *Cognitive Psychology*, 34, 72–107.
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10, 437–442.
- Potter, M. C. (1975). Meaning in visual search. *Science*, 187, 965–966.
- Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning*, 2(5), 509–522.
- Potter, M. C., Staub, A., & O'Connor, D. H. (2004). Pictorial and conceptual representation of glimpsed pictures. *Journal of Experimental Psychology: Perception and Performance*, 30(3), 478–489.
- Rensink, R. A. (2000). Seeing, sensing, and scrutinizing. *Vision Research*, 40(10–12), 1469–1487 (Review).
- Rensink, R. A. (2002). Change detection. *Annual Review of Psychology*, 53, 245–277.
- Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science*, 8, 368–373.
- Shepard, R. N. (1967). Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior*, 6, 156–163.
- Standing, L. (1973). Learning 10,000 pictures. *Quarterly Journal of Experimental Psychology*, 25(2), 207–222.
- Tanaka, J. W., & Presnell, L. M. (1999). Color diagnosticity in object recognition. *Perception and Psychophysics*, 61(6), 1140–1153.
- Torralba, A., Oliva, A., Castelhana, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. *Psychological Review*, 113(4), 766–786.
- Treisman, A. (1986). Properties, parts, and objects. In K. R. Boff, L. Kaufmann, & J. P. Thomas (Eds.), *Handbook of human perception and performance*. New York: Wiley and Sons.
- Wichmann, F. A., Sharpe, L. T., & Gegenfurtner, K. R. (2002). The contributions of color to recognition memory for natural scenes. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 28(3), 509–520.