Cognitive Psychology **41,** 176–210 (2000)
doi:10.1006/cogp.1999.0728, available online at http://www.idealibrary.com on IDEAL®

Diagnostic Colors Mediate Scene Recognition

Aude Oliva and Philippe G. Schyns

University of Glasgow, Glasgow, United Kingdom

In this research, we aim to ground scene recognition on information other than the identity of component objects. Specifically we seek to understand the structure of color cues that allows the express recognition of scene gists. Using the *L*a*b** color space we examined the conditions under which chromatic cues concur with brightness to allow a viewer to recognize scenes at a glance. Using different methods, Experiments 1 and 2 tested the hypothesis that colors do contribute when they are diagnostic (i.e., predictive) of a scene category. Experiment 3 examined the structure of colored cues at different spatial scales that are responsible for the effects of color diagnosticity reported in Experiments 1 and 2. Together, the results suggest that colored blobs at a coarse spatial scale concur with luminance cues to form the relevant spatial layout that mediates express scene recognition.

Key Words: scene; color; diagnostic information; recognition; categorization; spatial scale; L*a*b*; spatial layout.

In Potter's (1975) classical scene-recognition experiment, subjects faced a screen on which slides of real-world scenes appeared in rapid succession (at a rate of 125 ms/slide). Their task was to press a button as soon as they detected, e.g., a beach. Subjects' efficiency was very high and this presents a puzzling problem for scene analysis: how can a scene be so rapidly recognized despite its variability, large number of component objects, and multiple sources of interfering factors?

Following Marr's (1982) influential conception, scene recognition has

This work was partially funded by a Fyssen Foundation research fellowship (Paris, France) awarded to Aude Oliva and by a ESRC Grant R000237412 awarded to Philippe G. Schyns. The first experiment was presented at the European Conference of Visual Processing, Strasbourg, 1996. Matlab and C code to perform the color transformations and comparisons reported in this paper can be obtained by writing to Philippe G. Schyns.

Correspondence and reprint requests may be sent to either author: Aude Oliva, Center for Ophthalmic Research, Brigham and Women's Hospital, 221 Longwood Ave., Boston, MA 02115. Fax: (617) 277-2085. E-mail: oliva@warcn.bwh.harvard.edu. Philippe G. Schyns, Department of Psychology, University of Glasgow, 56 Hillhead Street, Glasgow G63 9RE, UK. Fax: +44 141 339 88 89. E-mail: philippe@psy.gla.ac.uk.

The authors thank Gregory L. Murphy, Harry C. Potter, and three anonymous reviewers whose comments greatly helped improve this manuscript.

0010-0285/00 $35.00

been portrayed as a progressive reconstruction of input from simple local measurements. Boundary edges, surface markers, and other visual cues are progressively integrated into successive layers of increasingly complex representations, the last of which assembles the identity of the scene from the identity of a few component objects. However, this depiction does not account for all situations of recognition. Scenes are often recognized very quickly, in a single glance—in fact, as fast as a single component object (Biederman, Mezzanotte, & Rabinowitz, 1982; Friedman, 1979; Intraub, 1997; Potter, 1976). Under these conditions, it has been shown that visual information could mediate scene recognition without the prior and necessary recognition of component objects (Schyns & Oliva, 1994).

The aim of the present paper is to ground scene recognition on information other than the identity of component objects. Specifically, we seek to understand the structure of the visual information that allows the ''express recognition'' of naturalistic, real-world scenes. Work in early vision suggests that the early bases of recognition information are the dimensions of luminance (brightness), chromaticity (color), and movement and depth (Livingstone & Hubel, 1987). In this paper, we examine the contribution of chromatic cues to scene recognition at a glance (see Oliva & Schyns, 1997; Schyns & Oliva, 1994, for the role of luminance cues). We first review the role of color in recognition before turning to three studies that examined the conditions of use and the structure of color cues for scene recognition.

Luminance, Color, and Recognition

Psychophysical research has revealed that early vision operates simultaneously with luminance and chromatic information in processing motion (Cavanagh & Ramachandran, 1988), texture (McIlhaga, Hine, Cole, & Schneider, 1990), stereo vision (Logothetis, Schiller, Charles, & Hurlbert, 1990), and simple shapes (Cavanagh, 1996; Damasio, Yamada, Damasio, Corbett, & McKee, 1980). In higher-level vision, numerous studies have examined how luminance cues supported the recognition of faces (Breitmeyer, 1984; Costen, Parker, & Craw, 1994; Fiorentini, Maffei, & Sandini, 1983; Schyns & Oliva, 1997, 1999; Sergent, 1986), objects (Parker et al., 1996), and scenes (Oliva & Schyns, 1997; Parker, Lishman, & Hughes, 1992; Schyns & Oliva, 1994). It was found that fine-scale boundary edges (from high spatial frequencies) and coarser scale blobs (from low spatial frequencies) could selectively mediate different categorizations of the same stimuli (e.g., Oliva & Schyns, 1997; Schyns & Oliva, 1999).

In marked contrast, little is known about the role of chromatic information in scene-recognition tasks—in fact, it is not even clear that the color dimension contributes at all to scene recognition, and so there is little research on the structure of color cues. Turning to object-recognition studies for a possible explanation, it appears that the role of color is controversial. Color is typically studied with two types of tasks: verification and naming. In verification, one lexical name is typically presented before an object picture. The name is supposed to activate a representation in memory to match against the input. When the activated representation comprises color cues, if these participate in recognition, then matching should be better with colored than with achromatic versions of the same object (Biederman & Ju, 1988; Joseph & Proffitt, 1996; Sanocki, Bowyer, Heath, & Sarkar, 1998). Studies of recognition at the basic level (Rosch, Mervis, Johnson, Grey, & Boyes-Braem, 1976) revealed that subjects verified pictures of common objects equally fast whether they were colored or not (Biederman & Ju, 1988; Davidoff & Ostergaard, 1988; Ostergaard & Davidoff, 1985). Consistent with the idea that colors play no role in bootstrapping recognition, Ostergaard and Davidoff (1985) found that objects were verified equally fast, irrespective of whether they were properly colored. Colors, however, should at least inform the recognition of some objects. For example, orange, the color, is highly predictive (i.e., diagnostic) of an *orange*, the fruit, and could therefore facilitate its recognition. Biederman and Ju (1988) showed that this was not even the case: objects judged to be high in color diagnosticity were verified no faster when colored than when black-and-white line drawings.

Object naming is the second task commonly used in studying the role of color. In naming, a picture is shown and subjects tag a lexical name from a restricted set of possibilities. Naming tasks produced markedly different results. For example, Davidoff and Ostergaard (1988) found that objects like fruits and vegetables were named faster when presented in color than when not (see also Davidoff & Ostergaard, 1988; Price & Humphreys, 1988; Wurm, Legge, Isenberg, & Luekber, 1993). Colors have also been found to affect the categorization of objects at levels higher than basic. Price and Humphreys (1989) found that naming an object as a ‘‘fruit’’ or a ‘‘vegetable’’ was faster when it was properly colored.

Tanaka and Presnell (1999) explicitly addressed the puzzling discrepancy between the verification and naming performances reported in the object-recognition literature. They conducted experiments that applied the two tasks to a set of common objects.[1] However, they controlled, using a feature-listing task, the use of diagnostic colors in each object category. In the feature-listing task, subjects saw pictures of objects (e.g., biological and human-made) and were instructed to quickly write down their three main perceptual features. For example, the picture of an orange could elicit the features ‘‘round,’’ ‘‘orange,’’ and ‘‘bumpy.’’ Each object was then coded for the presence and rank of its colors in the feature listings. With this control, the authors found faster verification *and* naming times for objects with higher color-diagnosticity rankings.

[1] It is interesting to note that most of these objects were used in other main studies of color in object recognition (e.g., Biederman & Ju, 1988; Davidoff & Ostergaard, 1985; Ostergaard & Davidoff, 1988; Price & Humphreys, 1989).

In summary of the reviewed evidence, it appears that color diagnosticity must be carefully controlled in recognition studies because diagnostic object colors can contribute to recognition performance. It is one goal of this research to develop a more rigorous control of color diagnosticity than feature listings. To the extent that a scene comprises several colored objects, its recognition speed could also benefit from colors when these are diagnostic of the objects themselves. However, we already pointed out that scenes can also be identified from scene-specific cues, not from the identity of their objects (Biederman, 1981; Henderson, 1992; Intraub, 1997; Sanocki & Epstein, 1997; Oliva & Schyns, 1997; Schyns & Oliva, 1994). Hence, a simple generalization from object to scene colors might be audacious. Another goal of the research is to determine the nature of the colored information that can facilitate fast scene recognition.

There is still little psychological research on the recognition of complex scenes *per se*. When scenes are used, it is to assess the role of perceptual context in object recognition (e.g., Aginski & Tarr, 2000; Biederman et al., 1982; Boyce, Pollatsek, & Rayner, 1989; Delorme, Fabre-Thorpe, Richard, Fize, & Thorpe, 1998; Hollingworth & Henderson, 1998; Intraub, 1997; Sanocki, Bowyer, Heath, & Sarkar, 1998). For example, Delorme et al. (1998) presented humans and monkeys 400 pictures of food and animals in their natural background scene contexts. Half of the pictures were colored; the other half were gray levels. Under tachistoscopic (20–30 ms) conditions of presentation and speeded judgments (around 400 ms in a go/no-go detection task), color affected slower responses only in the food category. On this basis, the authors argued for comparatively late effects of color in recognition. In contrast, Gegenfurtner (1998) showed that tachistoscopic (30–50 ms) presentations of colored pictures elicited better retrieval from memory than their luminance counterparts, illustrating that chromatic cues can index scene memory. This color advantage occurred irrespective of whether the target represented a human-made (e.g., streets) or a natural scene (e.g., flowers and landscapes). However, a scene-recognition experiment of Oliva and Schyns (1996) revealed a diagnostic influence of color. They compared the naming speed of briefly presented (30 or 120 ms) natural, color-diagnostic (e.g., *beach, forest, valley*) and artifact, color-nondiagnostic (e.g., *city, road, room*) pictures of real scenes. A panel of independent judges rated the color diagnosticity of the scenes. Under these conditions, natural scenes were recognized faster than their luminance counterparts, but artifacts were named equally fast. Existing data with real pictures therefore suggest that the color is never, always, and sometimes used to recognize a scene!

Methodological Considerations

Conflicting results in scene recognition and the reported controversy in object recognition stress the importance of carefully controlling both the diagnosticity of colors and the comparable nature of the stimuli (e.g., luminance vs normally colored vs abnormally colored) before making any general statement. To date, the literature on ''colors in recognition'' has not always acknowledged the importance of these controls and it remains difficult to assess whether color does or does not contribute to a speeded recognition task.

This research introduces the $L^*a^*b^*$ color space to human recognition experiments (see Appendix 1 for a formal presentation). $L^*a^*b^*$ will be used in all our experiments to better control the diagnosticity of scene colors and the conditions under which stimuli are visually presented. Three dimensions are sufficient to represent all colors of the visible spectrum. Different three-dimensional spaces are available to represent these colors (see Appendix 1). $L^*a^*b^*$ has a number of interesting properties: First, it explicitly separates luminance (L^*) on a first dimension from chroma (a^*b^*) on the two remaining dimensions. This enables a formal transformation of colors which has little effect on luminance information. Second, a^*b^* represents colors along two color-opponent dimensions: a^* extends from green to red and b^* from blue to yellow. It is generally agreed that the visual system processes color along such oppositions (see Logothetis, Schiller, Charles, & Hurlbert, 1989). The second interesting property of $L^*a^*b^*$ is that it is close to perceptual uniformity. In a perceptually uniform encoding, the Euclidian distance between any two points [$L1^*$ $a1^*$ $b1^*$] and [$L2^*$ $a2^*$ $b2^*$] mirrors the perceived difference of the colors they represent (this property holds for medium to long distances in $L^*a^*b^*$; see Wyszecki & Stiles, 1982, for a review and discussions). Connelly (1996) tested several color spaces (including $L^*a^*b^*$, $L^*u^*v^*$, and XYZ) to determine how robust they were to changes in chroma (color saturation) and hue. Using digitized pixels sampled from differently colored cardboards of the Munsell scale, she observed that $L^*a^*b^*$ encodings produced cleaner separations which preserved the topography of perceptual distances in humans (as expected by the property of perceptual uniformity of $L^*a^*b^*$).

The present research capitalized on these properties of $L^*a^*b^*$ to generate a new and powerful method of synthesizing photographic stimuli for which (1) a selective change of color does not affect physical luminance and (2) color diagnosticity can be controlled. Note that this is a significant departure from standard methods. The common RGB encoding does not separate luminance from chromaticity, and so a change in color produces a change in luminance. In $L^*a^*b^*$, however, we can independently change the a^* and b^* axes and recompose a new image whose colors have changed, but whose physical luminance has not. To illustrate, if a^* represents the green-to-red spectrum and b^* represents the blue-to-yellow spectrum, a swap of two axes ($L^*b^*a^*$) would change the color of a beach from yellow to red (see Fig. 1B). Swap is the first operator that can be applied to the color information of a scene in $L^*a^*b^*$. An inversion of values along the a^* and b^* axes is the second operation. For example, an inversion of b^* would create a blue beach. Of course, the two operators are independent, and *swap+invert* could synthesize a green beach (see Fig. 1C). In our experiments, we used composi-


A
B
C

tions of the swap and invert operators to produce differently colored versions of the same scene in which luminance was held constant (examples of stimuli are presented in Experiment 1).[2] This property of $L^*a^*b^*$ allowed better control of the typical conditions of stimulation of normally colored, abnormally colored, and luminance-only.

We used the perceptual uniformity of $L^*a^*b^*$ to control the diagnosticity of colors. Each picture was composed of a fixed number of pixels which projected into the two-dimensional, color-opponent a^*b^* space. When scenes had typical colors (e.g., the variations of yellows and browns of deserts), the colors of their pixels formed tight, separable clusters in a^*b^* (see Fig. 2). When they did not have distinct colors, their pixels overlapped in a^*b^* (see Fig. 3). We could therefore select scene categories with the constraint that their projections formed distinct, nonoverlapping, and roughly equidistant clusters in a^*b^*. The computation of this control is detailed in Appendix 2.

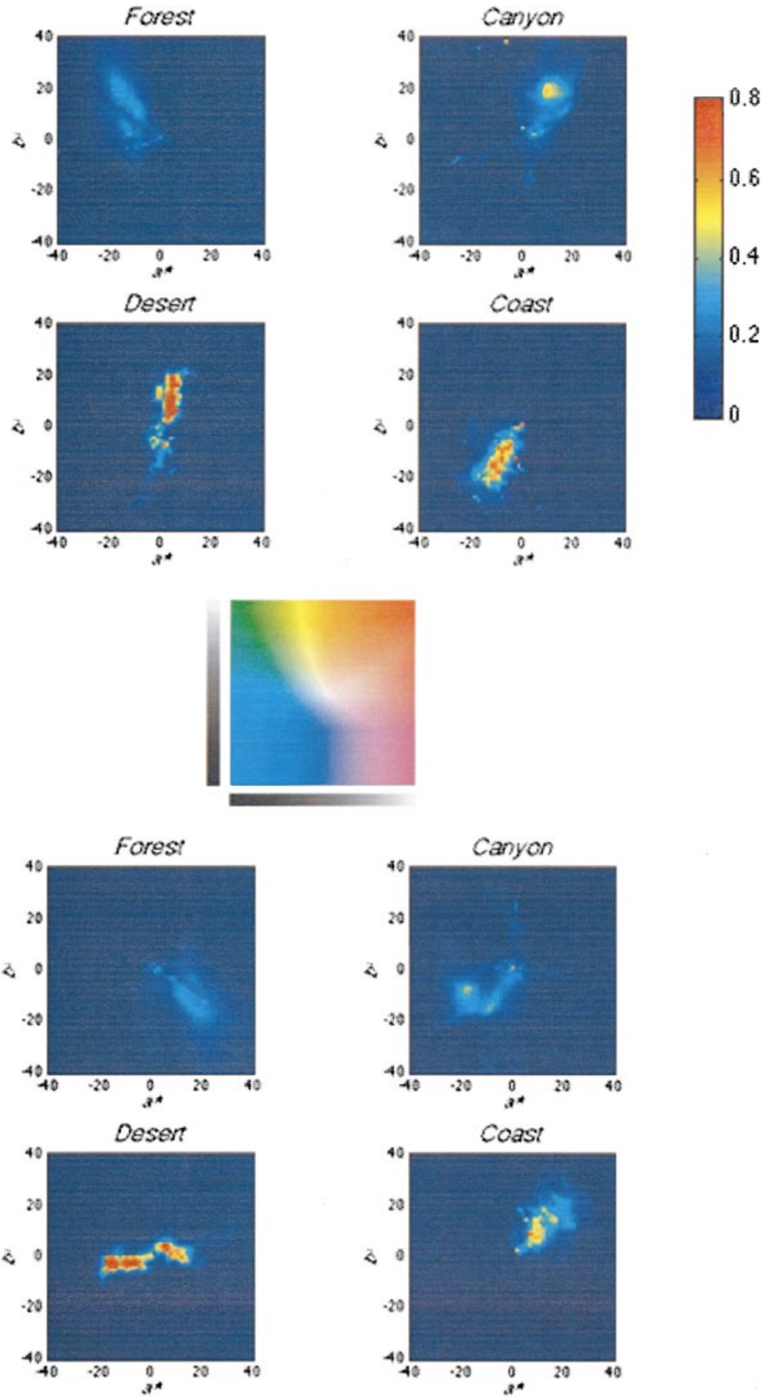
Three experiments which manipulated and controlled the diagnosticity of

---

FIG. 1. (B) The application of the operator axis_swap ($L^*a^*b^* \rightarrow L^*b^*a^*$) to the top beach picture: every pixel of the green/red opposition is transposed in blue/yellow and vice versa. Note, for example, that the yellow sand becomes redish and that the green sea becomes blue. The bottom picture illustrates the application of the axis inversion operator to the ''axis swapped'' middle image: each red pixel (e.g., the sand in the middle picture) becomes greenish and each blue pixel (e.g., the sea in the middle picture) become yellow in the bottom picture. Applications of axis swap and inversion were used to synthesize the abnormally colored scenes of our experiments.

FIGS. 2 and 3. These figures illustrate the color histograms we used to control the color diagnosticity of scenes in Experiment 1. To compute each histogram, we divided the a^*b^* space in a lattice of 80×80 equally spaced bins. For each category, we retrieved the a^*b^* bin of each pixel of each exemplar by transforming it from RGB to $L^*a^*b^*$. We then normalized the outcome by summing the pixels in each bin and divided the total per bin by the total number of pixels per category. The axes of each histogram represent the coordinates of each color in the 80×80 lattice covering a^*b^*. The frequency of each color is represented in a percentage. The high resolution of the histograms produced percentages typically varying between 0 and 1 for each of the 80×80 possible color bins. Increasing color intensities in a ''rainbow encoding'' (see the color bar in Fig. 2) mirror increasing percentages. This technique produced a different color histogram per category in a^*b^*. The color histograms of Fig. 2 do not overlap, indicating that the categories were color-diagnostic whereas those of Fig. 3 do overlap, revealing that the categories were not color-diagnostic. The upper four histograms of Figs. 2 and 3 show the color histograms of the normally colored scenes, whereas the lower four show the abnormally colored version. The depiction of a^*b^* in the center helps to understand the main colors of each histogram.

[2] The images used in our experiments had an average luminance of 128 ($STD = 70$) on a [1 . . . 256] gray-level scale. The transformation of color could sometimes introduce a marginal change so that the average luminance level would oscillate between 124 and 132.



Forest
Canyon
Desert
Coast
0.8
0.6
0.4
0.2
0
40
20
0
-20
-40
a*
Forest
Canyon
Desert
Coast

FIG. 2.

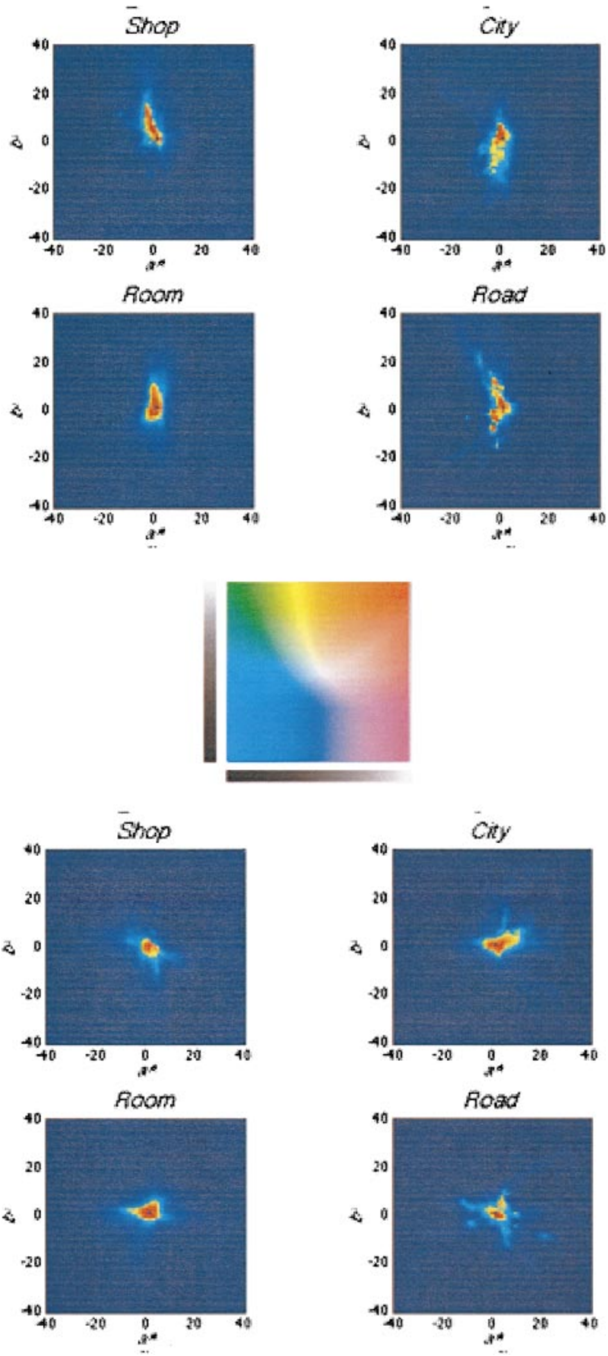


FIG. 3.

scene colors in $L^*a^*b^*$ were run to examine their impact on scene-recognition performance. Remember that the object-recognition literature reported different effects with verification and naming tasks (Biederman & Ju, 1988; Davidoff & Ostergaard, 1988; Price & Humphreys, 1988; Ostergaard & Davidoff, 1985; Wurm, Legge, Isenberg, & Luekber, 1993; but see Tanaka & Presnell, 1999). Experiments 1 and 2 therefore tested whether diagnostic colors influence scene naming and verification, respectively. Experiment 3 addressed the issue of the nature of luminance and color cues in scale space and their respective contributions to scene recognition.

EXPERIMENT 1

Experiment 1 tested the hypothesis that colors contribute to recognition when they are diagnostic of a scene category. Remember that $L^*a^*b^*$ separates luminance (L^*) from two independent chromatic oppositions (a^*, red–green, and b^*, blue–yellow). We chose four categories of scenic photographs (*canyon*, *forest*, *seashore*, and *desert*) with the constraint that their main colors stood at the opposite sides of the a^*b^* axes. The top four pictures of Fig. 2 illustrate that the projections of these categories in a^*b^* (i.e., the pixels of each of their exemplars) occupy four distinct, nonoverlapping, and roughly equidistant regions. Canyons are mostly red and orange, forests green, seashores blue, and deserts yellow (compare with the depiction of a^*b^* provided in the picture). Chromatic information is therefore objectively diagnostic of these categories in the task considered. In contrast, the projection of the four other categories (*city*, *shops*, *road*, and *room*) do not form tight, nonoverlapping clusters in a^*b^*—i.e., they all occupy a similar area of the space (see the top four pictures of Fig. 3). Color cannot be diagnostic of these categories.[3]

The opposition between color-diagnostic and color-nondiagnostic scenes is the backbone of Experiment 1: we expected faster recognition when colors were diagnostic—i.e., the scenes represented in Fig. 2. There is one potential confound with this approach. We aim to assess the role of diagnostic colors on recognition, but the task allows a strategy independent of proper scene recognition: subjects could learn to respond to diagnostic colors without necessarily recognizing the scenes themselves. Such an effect of colors in the task, not of colors on recognition per se, must be controlled. To this end, we synthesized abnormally colored versions of the color-diagnostic categories with the constraint that they projected onto other well-separated regions of a^*b^* as did their normally colored versions. The lower pictures of Fig. 3 illustrate the desired result that abnormally colored categories were still

[3] This is not to say that a given living room does not have specific colors, but that *living room*, as a category of different pictures, does not form a cluster distinct from the other categories chosen.

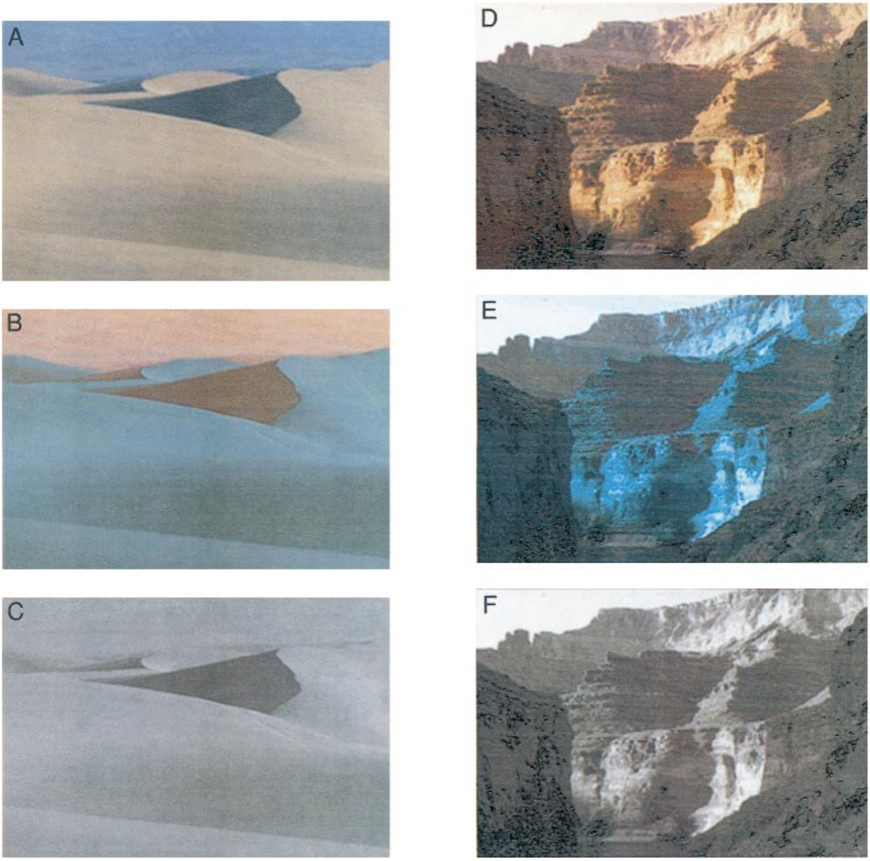


FIG. 4. The three versions of two of the scene pictures used in Experiment 1. (A and D) Normally colored scenes (Norm). (B and E) Abnormally (Abn) colored versions. (C and F) Luminance-only (Lum) versions.

diagnostic in the task (they had distinct modal colors), but not in the real world (see Fig. 4 for examples of these scenes). Abnormally colored versions were also generated for the other scenes, but for reasons to be later explained, they could not be diagnostic in the task (see the bottom four histograms of Fig. 3).

In sum, Experiment 1 comprised three different versions of each color-diagnostic and nondiagnostic scenes: normally colored, abnormally colored, and luminance-only (or gray levels, see Fig. 4). The latter served to assess baseline recognition in the absence of any chromatic cues. We predicted differential effects of colors on recognition: color-nondiagnostic scenes should be recognized equally fast irrespective of whether they are colored; color-diagnostic scenes should be recognized faster when properly colored.

Methods

Participants

Sixteen Glasgow University students (between 18 and 25 years of age with normal or corrected vision) were paid to participate in the experiment.

Stimuli

The image sample comprised 160 color pictures of scenes selected from the Corel CD Photo Library. Images were divided into four categories of color-diagnostic scenes (*canyon, forest, coastline,* and *desert*) and four categories of color-nondiagnostic scenes (*city, shopping area, road,* and *room*). Twenty exemplars per category represented the scenes from a variety of viewpoints and perspectives. Note that the pictures had a similar luminance—i.e., their gray-scale means and standard deviations were similar.

Computation of color diagnosticity in $L^*a^*b^*$. Color-diagnostic categories were chosen with the constraint that their pixels projected into clearly distinguishable clusters in a^*b^*. Color-nondiagnostic categories were chosen so as not to form clear-cut clusters in a^*b^* (see Appendix 2 for the details of the computation).

Computation of conditions of stimulation in $L^*a^*b^*$. The experiment used three different versions of each scene: normally colored (Norm), luminance only (Lum), and abormally colored (Abn). The Norm scenes (160 pictures) were simply the original pictures. Lum stimuli discarded the chromatic components a^*b^* of Norm scenes, leaving only L^*, the gray levels. To compute Abn stimuli, we transformed the colors of Norm scenes in a^*b^*. Remember that axis_swap in a^*b^* replaces red–green pixels with blue–yellow pixels and vice versa. Axis inversion produces the opposite of a color (e.g., green pixels become red or vice versa; blue pixels become yellow or vice versa). We applied these transformations to the original scenes with the constraints of (1) producing tight clusters in locations of a^*b^* different from those of the normally colored pictures and (2) spanning about the same colors as those of the original stimuli. This could only apply to color-diagnostic scenes, as a distinct color mode is needed to produce another mode with the color transformation methods described here. Figure 4 shows two examples of the three types of stimuli. Color-nondiagnostic scenes were also color transformed but this did not change the diagnosticity of their colors. The color histograms of color-transformed scenes are shown in the lower pictures of Figs. 2 and 3.

Stimuli were presented on a Macintosh computer monitor. They were 472×325-pixel images, presented at a 150-cm viewing distance to subtend 6.4×4.4 degrees of visual angle. From 160 original pictures, we synthesized a total of 480 experimental stimuli equally divided between Norm, Abn, and Lum.

Procedure

The experiment comprised 480 trials: 160 Norm, 160 Abn, and 160 Lum. A trial started with a 500-ms presentation of a mid-gray fixation square followed, 100 ms later, by a picture (Norm, Abn, or Lum) displayed for 120 ms. Subjects were instructed to name aloud the image as quickly and as accurately as they possibly could. A 1500-ms latency separated two trials. Trials were divided into 10 blocks of 48, with within- and between-block randomization of order of presentation across subjects. Subjects were allowed a 1-min pause between blocks.

We instructed subjects that the scenes could only belong to one of eight possible categories (the names were explicitly listed and subjects were told to only use one of them). A vocal key connected to the computer measured the latencies between stimulus presentation and category name. Subjects were not shown the stimuli prior to the experiment. Eight practice trials were used to calibrate the vocal key and to familiarize subjects with the apparatus. The experiment lasted for about 50 min.

Results and Discussion

Error Rates

The results of one subject were discarded from the analysis because of abnormally high categorization errors (over 40%). The error rates of the remaining subjects were low (*city* = 5.2%; *road* = 4.2%; *room* = 4.2%; *shop* = 5.7%; *canyon* = 5.8%; *coast* = 5.6%; *desert* = 4.9%; *forest* = 3%), and did not differ across types of stimuli (Norm = 4.9%, Abn = 4.7%, Lum = 4.9%, respectively). A two-way ANOVA (eight categories × three stimulus types) on percentage correct categorizations did not reveal a significant difference between categories, [$F(7, 98) = 1.65$, *ns*] and type of stimuli [$F(2, 28) < 1$, *ns*]. Subjects were therefore able to correctly categorize all stimuli, and we can turn to an analysis of their reaction times.

Reaction Times

In this analysis we used only the RTs of correctly categorized trials with latencies within 2.5 *SD* from the mean. A two-way, within-subject ANOVA with type of category (color-diagnostic vs -nondiagnostic) and stimulus conditions (Norm, Abn, and Lum) revealed a main effect of diagnosticity [$F(1, 14) = 32.6$, $p < .0001$], a main effect of stimulus condition [$F(2, 28) = 21.98$, $p < .0001$], and a significant interaction [$F(2, 28) = 21.82$, $p < .0001$; see Fig. 5].

Figure 5 illustrates that chromatic information did indeed influence recognition in a scene-naming task. However, the interaction circumscribes the influence to the four color-diagnostic categories. A post hoc Tukey test revealed that normally colored stimuli were recognized more quickly than the same scenes without their colors [Lum = 804 ms, Norm = 758 ms, $F(1, 28) = 51.91$, $p < .0001$]. In contrast, these versions of color-nondiagnostic scenes were recognized equally as fast [Lum = 755 ms, Norm = 746 ms, $F(1, 28) < 1$, *ns*]. This confirms that the addition of colors to luminance cues in the color-diagnostic categories speeded their recognition.

The issue of the origin of the effect of color diagnosticity (task or prior knowledge) can now be addressed. Remember that abnormally colored scenes were constructed to be diagnostic in the task, but not in the real world (whereas the normally colored versions were diagnostic in the task *and* in the real world). Figure 5 shows that these scenes were in fact recognized more slowly than their achromatic, luminance-only counterparts [Lum = 804 ms, Abn = 824 ms, $F(1, 28) = 9.78$, $p < .02$]. Such interference was not observed for color-nondiagnostic scenes, which were again recognized equally fast in both conditions [Lum = 756 ms, Abn = 753 ms].

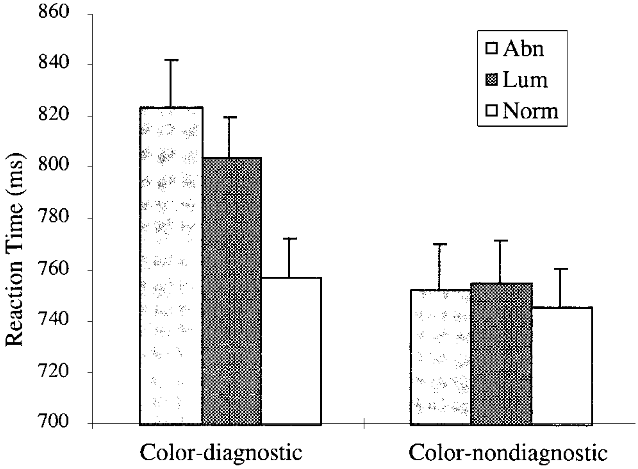


FIG. 5. Subjects' naming reaction times in the within-subjects design of Experiment 1. Performance was very similar across luminance-only, normally colored, and abnormally colored versions of the color-nondiagnostic scenes. In contrast, facilitation of normally colored pictures and intereference of abnormally colored pictures were observed for color-diagnostic scenes.

In sum, it appears that color influences recognition when it is diagnostic of a scene category: The addition of normal colors to an achromatic stimulus accelerates its recognition, whereas the addition of abnormal colors impedes it. At this stage, it is worth pointing out that these trends were observed for all four color-diagnostic categories (see Fig. 6). However, nondiagnostic colors had no observable effect on categorization speed. Appendix 3 presents a replication of all these effects of diagnosticity with shorter, 30-ms presentations of the stimuli. Thus, there is no doubt that diagnostic colors have selective influences on the naming of scenes, even when they are very briefly seen.

EXPERIMENT 2

Remember that object-recognition studies revealed a discrepancy between naming tasks, which produced effects of colors, and verification tasks, which did not. The naming task of Experiment 1 also produced effects of color. Experiment 2 therefore had two main aims: (1) to replicate in a verification task, and using a wider range of categories, the diagnostic effects of colors reported in Experiment 1 and (2) to examine how the luminance and chromatic cues memorized to represent a scene can contribute to recognition.

In a category-verification task a name precedes a picture and subjects must assess whether the two match.[4] In Experiment 2, the name was one of eight

[4] Verification tasks require ''yes/no'' responses and therefore authorize the use of many categories while limiting the interference of number of possible responses on reaction times.

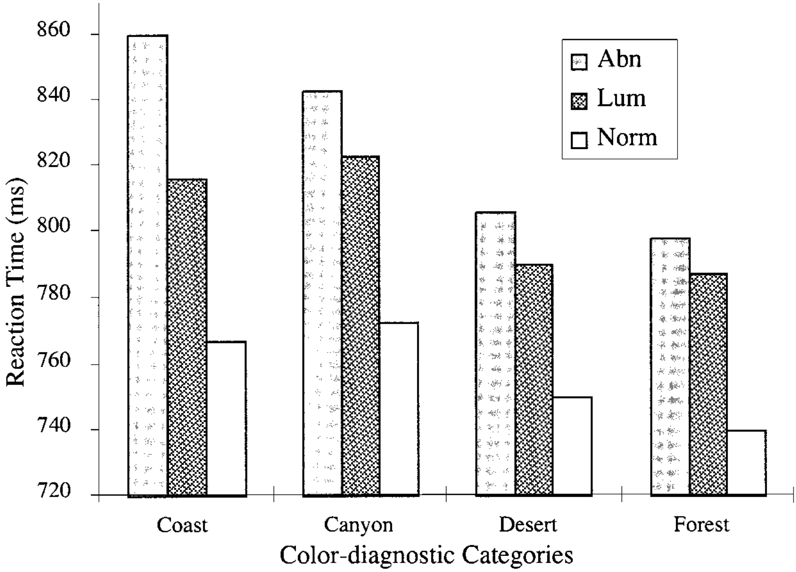


FIG. 6. The decomposition of the reaction times for each color-diagnostic category of Experiment 1. The trends of facilitation of normal colors and interference of abnormal colors applied to each of the tested categories.

possible color-diagnostic categories (*beach, canyon, coast, desert, field, forest, garden,* and *valley*), or eight color-nondiagnostic categories (*bathroom, bedroom, kitchen, living-room, city, restaurant, road,* and *shop*). Scenes were normally and abnormally colored. We measured the time subjects took to verify whether the name and the exemplar matched.

Experiment 2 comprised two types of trials: positive and negative. In positive trials, name and picture matched (e.g., ''beach'' followed by a beach picture). In agreement with the results of Experiment 1, we expected that subjects would match the category name faster with a normally colored scene than with an abnormally colored scene, but only when color was diagnostic of the category. When color was not diagnostic, we expected verifications to be equally fast for normally and abnormally colored scene exemplars.

In negative trials, name and picture did not match. This is a more complex situation because luminance and chromatic cues can differ in their mismatch with the named category, but it can also reveal an interesting trend. For example, in analogy to the Stroop effect, it should be comparatively easier to say ''no'' when diagnostic colors differ between the name and the target (as in ''field'' and desert) than when they match (as in ''garden'' and forest). Note that this prediction runs against most object-recognition studies which did not report effects of colors in verification tasks (though see Tanaka & Presnell, 1999). A subset of negative trials was designed to separate the respective contributions of memorized luminance and chromatic cues to the verification of color diagnostic categories.

In sum, the positive trials of Experiment 2 were designed to replicate the diagnostic effect of color in a verification task. A subset of negative trials tested the possibility of a Stroop interference.

Methods

Participants

Twenty University of Glasgow students with normal or corrected vision (age group, 18–25) were paid to participate in the experiment.

Stimuli

We used 128 colored scenes from the Corel CD Photo Library. Each belonged to one of 16 possible categories (with 8 different pictures per category), equally distributed between 8 color-diagnostic (*beach, canyon, coast, desert, field, forest, garden, and valley,* see Fig. A2.1 in Appendix 2) and 8 color-nondiagnostic categories (*bathroom, bedroom, kitchen, living room, city, restaurant, road,* and *shopping area;* see Fig. A2.2 in Appendix 2).

Computation of color diagnosticity in $L^*a^*b^*$. As in Experiment 1, we computed color histograms to assess that color-diagnostic categories had distinct colors in $L^*a^*b^*$ whereas color-nondiagnostic categories overlapped (see Appendix 2).

Computation of conditions of stimulation in $L^*a^*b^*$. Each scene picture came in two versions, normally (Norm) and abnormally (Abn) colored, computed in $L^*a^*b^*$ as explained previously. In total, Experiment 2 comprised 256 stimuli—128 Norm + 128 Abn = 256 divided into 8 exemplars/category × (8 Nat + 8 Art) × (Norm + Abn).

Positive Trials

In positive trials, the category name and the target picture belonged to the same category. They formed a total of 256 trials, testing all 16 categories with normally and abnormally colored scenes.

Negative Trials

In negative trials, name and picture belonged to different categories. A total of 384 trials tested all 16 categories using normally and abnormally colored versions of the scenes. Negative trials comprised 3 sets: 128 using color-diagnostic scenes, 128 using color-nondiagnostic scenes, and 128 using abnormally colored versions of color-diagnostic and color-nondiagnostic scenes. Of these trials, we were only interested in the first color-diagnostic set to test the Stroop interference. The others trials served as fillers.

The set of 128 color-diagnostic trials was further divided into 2 subsets of 64. The first 64 trials tested the four possible mismatches between a named category and luminance and chromatic cues: similar luminance and similar color (SL-SC), dissimilar luminance and similar color (DL-SC), similar luminance and dissimilar color (SL-DC), and dissimilar luminance and dissimilar color (DL-DC) (see Table 1). Here, similar color means highly correlated color histograms (e.g., garden and forest or valley and beach; see Table A2.3 in Appendix 2). Similar luminance denotes a similar gray-scale scene layout (e.g., a field and a desert are more similar in luminance than a field and a forest). A panel of four judges assigned the negative trials to the cells of Table 1 and the correlations between color histograms of color diagnostic categories were used to confirm the assignment (see the bold correlations in Table A2.3 in Appendix 2). Each trial of Table 5 was repeated 8 times, for a total of 64 negative trials. To establish

TABLE 1
Contingency Table of Negative Trials in Experiment 2

	Same structure	Different structure
Same color	SS-SC	DS-SC
	‘‘garden’’ → forest	‘‘beach’’ → valley
	‘‘forest’’ → garden	‘‘valley’’ → beach
Different color	SS-DC	DS-DC
	‘‘field’’ → desert	‘‘canyon’’ → coast
	‘‘desert’’ → field	‘‘coast’’ → canyon

baseline negative verifications, 64 negative trials randomized the category name presented before a color-diagnostic picture to destroy the luminance and chromatic associations tested in Table 1.

Procedure

A trial comprised one category name (*bathroom, beach, bedroom, canyon, city, coast, desert, field, forest, garden, kitchen, living room, restaurant, road, shopping area,* or *valley*) initially presented on the screen for 1000 ms, immediately followed by an 800-ms blank and presentation of a scene for 120 ms. Pictures subtended 6.4 $\times$ 4.4 degrees of visual angle at a 150-cm viewing distance. Subjects were instructed to decide as accurately and as quickly as they possibly could whether the name and picture matched by pressing one of two keys (half of the subjects used their right hand for the ‘‘yes’’ answer and the left hand for ‘‘no,’’ the other half did the opposite). A 1000-ms interval elapsed between two trials. The experiment comprised 640 trials split into 8 blocks of 80 trials. Presentation of blocks and trials within blocks were randomized. Subjects were allowed a 2-min pause between blocks. Prior to the experiment, subjects received a brief description of each scene category and its main components. They were asked to try to form a mental picture of the named scene. We hoped that this would mobilize all the visual dimensions of their scene representations.

Results and Discussion

Positive Trials

Subjects made very few verification errors: negative answers to positive trials occurred on less than 1% of all positive trials. We compared verification RTs of correct positive trials. A two-way, within-subjects ANOVA with categories (all 16 of them) and colors (Norm vs Abn) as factors revealed significant main effects of categories [$F(15, 285) = 17.33$, $p < .001$] and colors [$F(1, 19) = 33.41$, $p < .001$] and a significant interaction [$F(15, 285) = 4.114$, $p < .001$].

Analysis of the interaction did not reveal a significant difference between the verification RTs of normally and abnormally colored color-nondiagnostic scenes [Norm = 561 ms, Abn = 560 ms, $F(1, 285) < 1$, *ns*].[5] In contrast,

[5] The mean RTs, for Norm and Abn versions of the color-nondiagnostic categories, are (*SD* error in parentheses): *bathroom* 613–606 (11.2), *bedroom* 547–535 (9.3), *city* 476–475 (9.3), *kitchen* 559–564 (11.4), *restaurant* 568–560 (10.7), *road* 493–505 (11.6), *living-room* 620–634 (10.6), and *shopping area* 614–605 (11.3); emphasizing that colors did not play any role.

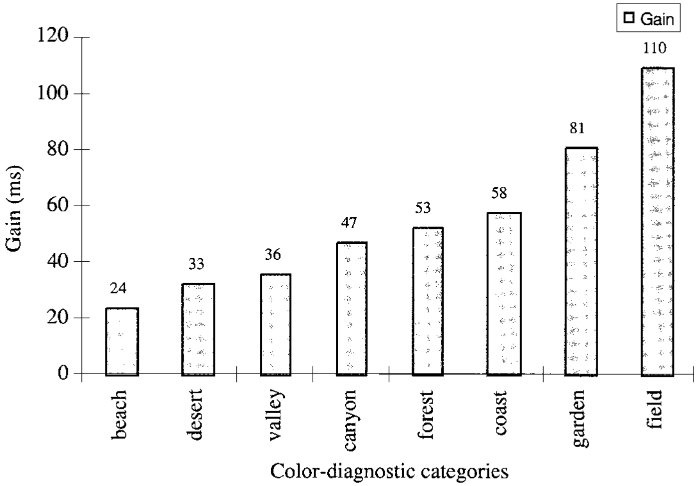


FIG. 7. The average gain in verification time as a function of color-diagnostic category in Experiment 2. The gain was computed by subtracting the average verification time of normally colored scenes from the average verification time of abnormally colored scenes.

such a difference existed for color-diagnostic scenes [Norm = 560 ms and Abn = 615 ms, $F(1, 285) = 85.04$, $p < .0001$]. Figure 7 illustrates that the advantage of normally colored scenes applied to all of the tested categories.[6] These results obtained with a verification task confirm with a larger set of categories the main effects of diagnostic colors observed in the naming task of Experiment 1. They indicate that color diagnosticity, not the experimental task, accounts for the reported advantage.

Negative Trials

Remember that a subset of the negative trials, those using color-diagnostic scenes, served to construct a Stroop interference paradigm. Subjects were asked to form a mental picture of the scene from the provided category name. We therefore expected that when colors matched between the name and the subsequent target picture, subjects would find it comparatively harder to establish that when they did not match. Such interference would further demonstrate that memorized diagnostic colors are actively used over the course of scene recognition.

Errors (i.e., saying ‘‘yes’’ when the name and target differed) were negligi-

[6] Note that the advantage is here expressed as the absolute difference between the reaction times to normally and abnormally colored scenes. Specific mean RTs for Norm and Abn were (*SD* in parentheses): *beach,* 607–631 (11.4), *canyon* 548–595 (10.8), *coast* 614- 672 (12), *desert* 526–559 (33), *field* 521–631 (11.9), *forest* 518–571 (12), *garden* 518–599 (11), and *valley* 631–667 (11.4). Note that there was no positive correlation between mean RT and gain (corr = −.25).

ble. We analyzed the RTs of correct negative trials using a three-way, within-subjects ANOVA, with color (similar vs dissimilar), luminance structure (similar vs dissimilar), and interference (baseline vs other trials) as factors. We found a significant main effect of color [$F(1, 19) = 13.58$, $p < .01$], no main effect of luminance [$F(1, 19) < 1$, *ns*], and a significant interference [related trials = 591 ms and baseline trials = 539 ms, $F(1, 19) = 141.53$, $p < .0001$]. The interaction color × interference was significant, [$F(1, 19) = 58.24$, $p < .0001$], whereas the interaction luminance x interference was not. Together, these results suggest that the diagnostic colors associated with a category name (e.g., ''garden'') interfered with the decision that the subsequently presented scene (e.g., *forest*) did not match.

This point was further explored with a decomposition of the double interaction, itself significant [$F(1, 19) = 5.31$, $p < .05$]. Interference on verification was stronger when category name and target picture were mismatched in both luminance and color [118 ms interference, SL-SC = 648 ms, compared to a baseline of 530 ms; $F(1, 19) = 147.22$, $p < .0001$]. A strong 80-ms interference also appeared when only colors were similar between name and picture [DL-SC = 605 ms, its baseline = 525 ms; $F(1, 19) = 67.27$, $p < .0001$]. In contrast, no interference arose when colors mismatched between name and picture, irrespective of the luminance similarities (i.e., the SL-DC = 547 ms, baseline = 544 ms, and DL-DC = 565 ms, baseline = 555 ms).

To summarize Experiment 2, positive trials revealed faster verifications of category membership of scenes when they were color-diagnostic and properly colored. Color-nondiagnostic categories were verified equally fast, whether scenes were properly colored or not. Negative trials with color-diagnostic categories revealed a Stroop effect when colors matched between a name and a scene from a different category. Together with the results of Experiment 1, we can conclude that color diagnosticity, not the experimental task itself (i.e., naming or verification), determines whether color exerts an influence on speeded-scene recognition tasks (see also Tanaka & Presnell, 1999, for a similar conclusion with objects).

EXPERIMENT 3

Experiments 1 and 2 have established that diagnostic colors participate in the recognition of real-world scenes. This raises the question of the nature of the chromatic cues tapping into the recognition mechanisms. Schyns and Oliva (1994; see also Oliva & Schyns, 1997) reported that coarse-scale luminance cues (around 2 cycles/degree of visual angle, corresponding to 8 cycles/image) could mediate the accurate classification of scenes into broad categories (e.g., *room, city, highway, mountain,* and so forth). At a coarse scale, however, a scene cannot be identified from its objects (see the most filtered scenes in Fig. 8) and scene specific configural information must mediate its recognition.

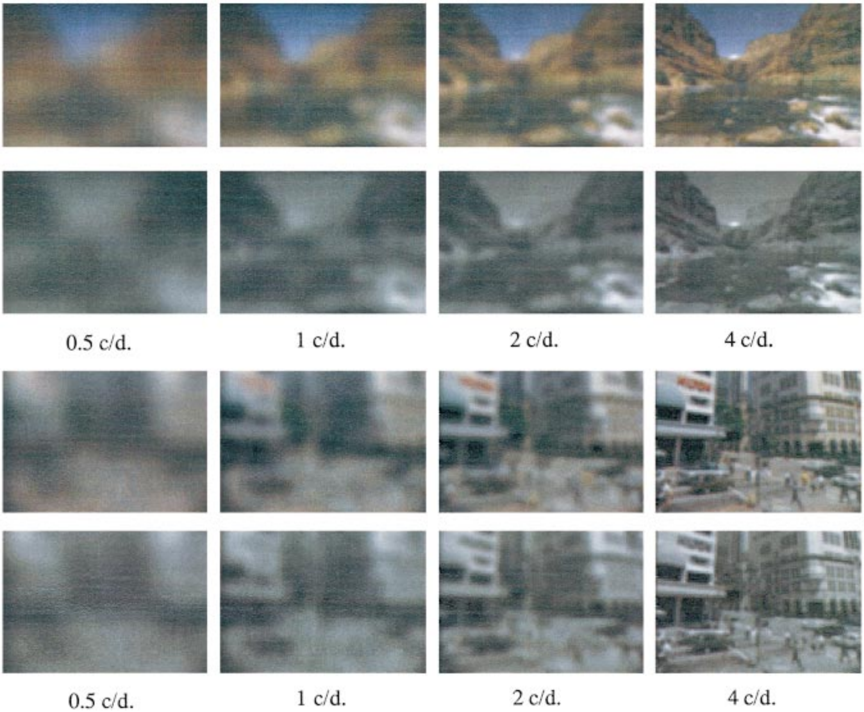


FIG. 8. The different conditions of stimulation used in the free-categorization task of Experiment 3 in the colored and luminance-only conditions. The upper scene is color-diagnostic, whereas the lower scene is not.

Even if it has been observed that pictures of natural scenes do have color information represented in medium to high spatial frequencies (Parraga et al., 1998), the visual system tends to perceive chromatic information at coarser scales better than luminance information (Mullen, 1985).[7] The reported effects of color should therefore arise from the spatial layout of crude color information.

Experiment 3 examined the hypothesis that the addition of colors to luminance cues improves the efficiency of scene categorization. To this end, we low-passed 24 colored scenes (12 color-diagnostic and 12 color-nondiagnostic) starting with a cutoff of 0.5 cycle/degree climbing progressively to 8 cycles/degree of visual angle (see the examples presented in Fig. 8). This produced a spectrum of scene information at different scales. One subject group categorized these scenes. To isolate the role of color, a second

[7] It has been observed that the contrast sensitivity for chromatic information drops off faster at high spatial frequency than luminance contrast. This perceptual constraint would suggest that useful color information would tend to be in the lower part of the spatial frequency spectrum.

group categorized the luminance-only versions of the same scenes. We could therefore measure the gains in categorization performance that arose from the addition of color cues to luminance information at different spatial scales.

On the basis of previous experiments, we expected categorization performance to be high (i.e., superior to 70% accuracy) for luminance-only images with a spatial frequency content above 2 cycles/degree of visual angle. The results of Experiments 1 and 2 also led us to predict that if colored blobs are useful categorization cues, then adding colors at different levels of spatial scales should offer a differential advantage to color-diagnostic and color-nondiagnostic scenes. Namely, scenes at very low resolutions should benefit more from the addition of colors when these are diagnostic of the category. Color-nondiagnostic scenes could also benefit from color, but not from very low resolutions (because the organization of colored blobs is not diagnostic of the category considered). Together, these results would establish that colored blobs can be the diagnostic information required for scene recognition.

Methods

Participants. Thirty University of Glasgow students (between 18 and 25 years of age with normal or corrected vision) were paid to participate in the experiment. They were randomly assigned to one of two experimental conditions.

Stimuli. We selected a total of 24 scenes, each from a different category, with the constraint that they could be individually named. The sample comprised 12 color-diagnostic scenes (*beach, canyon, coast, desert, field, forest, lake, mountain, port, river, countryside road,* and *valley*) and 12 color-nondiagnostic scenes (*bathroom, bedroom, kitchen, living-room, office, stairs, skyscraper, city center, highway, restaurant, shopping area,* and *street*).

Each scene was low-passed at six different levels, 0.5, 1, 1.5, 2, 4, and 8 cycles/degree of visual angle.[8] Filtering was independently performed in L^*, a^*, and b^*; luminance-only stimuli were L^* (see Fig. 8). In total, the experiment comprised 144 colored images and their 144 luminance counterparts.

Procedure. Participants sat 150 cm away from the computer screen so that each scene subtended 6.4×4.4 degrees of visual angle. Images were randomly presented one at a time for 150 ms. In a free naming task, subjects were instructed to name each scene they saw. Instructions emphasized that pictures were scenic photographs quickly displayed on the screen with some of them blurred, that the pictures only represented indoor and outdoor scenes, and that there were no close-up views of objects. Subjects were told to look at the entire picture to determine the displayed scene. Instructions also specified that identification should be as precise as possible, using only one or two words to categorize each scene (e.g., garage, classroom, garden, and so forth). Subjects were told not to use imprecise names such as ''indoor'' or ''outdoor.'' To minimize response biases, subjects did not see the actual pictures prior to the experiment and they were not told which category names to use. They were required to write down one answer per scene in the proper slot of a preformatted answer sheet without receiving any form of corrective feedback. A key-press on the computer keyboard presented the next picture. Completion of the experiment took about 1 h.

Results and Discussion

As subjects performed a free naming task, we coded their categorization responses. Three independent judges assessed whether the names given cor-

[8] These values corresponded to 1 cycle every 64 pixels of the image, for the lowest filtering, to 1 cycle every 4 pixels, for the highest level of filtering.

responded to the presented scenes. For example, a coast scene could be called a coast, a sea coast, a coastline, a seashore, a bay, or a view on the ocean, and the city scene of Fig. 8 was named a city, buildings, a street, a city center, and an urban zone. The scene was properly categorized whenever the name and picture were judged to be congruent. A different score (where 100% means 24 correct categorizations) was therefore computed for each tested spatial resolution of each scenic picture.

We excluded two participants (one per group) from the analysis because they did not follow the instructions and produced disproportionately high errors (more than 50%). A three-way ANOVA on percentage correct categorization with between-subjects factor ''luminance vs colored stimuli'', within-subjects factors ''color-diagnostic vs -nondiagnostic scenes'', and six levels of spatial filtering revealed a main effect of colors [51% vs 76% correct for luminance and colored, respectively; $F(1, 26) = 57.97$, $p < .001$], a main effect of filtering level [$F(5, 130) = 279.17$, $p < .001$], and no global significant difference between correct categorizations of color-diagnostic (64%) and -nondiagnostic (63%) scenes.

Figure 9 illustrates the significant interaction between color (the plain lines) vs gray scale (the dashed lines) and levels of filtering, [$F(5, 130) = 19.68$, $p < .001$].The addition of colors greatly enhanced performance at low resolutions, up to 1.5 cycles/degree, resulting in a shift of the performance curves corresponding to the colored stimuli (the plain lines). At 0.5 cycle/degree of filtering, correct categorizations were at 36% with colored scenes, but at only 2% (with chance 1/24 = 4.25%) for the same stimuli without colors. The level of performance at 1 cycle/degree with colored scenes was only attained one octave higher, at 2 cycles/degree, with their achromatic counterparts. Consequently, correct categorizations are mostly found with colored stimuli at coarse resolutions. Starting from 1.5 cycles/degree, correct categorizations become more evenly distributed between luminance-only and colored stimuli. These initial results illustrate that the addition of coarse chromatic cues to the luminance blobs of a scene clearly enhanced its categorization, but also that the information supporting this enhancement resided below 2 cycles/degree (or 8 cycles/image) of spatial resolution.

Turning to the selective influence of color diagnosticity on categorization performance, a significant interaction was found between this factor and levels of spatial filtering [$F(5, 130) = 9.26$, $p < .001$] as well as a double interaction [$F(5, 130) = 2.77$, $p < .05$]. A decomposition of the latter revealed that the former was significant for the categorizations of both luminance-only [$F(5, 130) = 8.42$, $p < .001$] and colored stimuli, $F(5, 130) = 3.62$, $p < .01$]. For colored stimuli, the early separation between the bold and thin lines in Fig. 9 indicates an influence of diagnostic colors on scene categorization—in fact, the difference is significant as early as 0.5 cycle/degree between color-diagnostic (49%) and color-nondiagnostic (23%) scenes [$F(1, 130) = 45.98$, $p < .001$]. Remember that the dashed

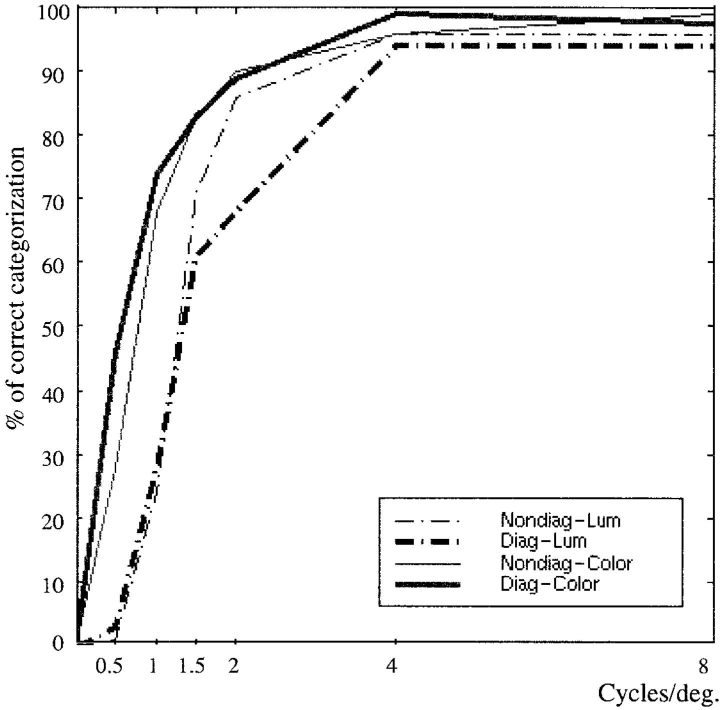


FIG. 9. The response curve of subjects submitted to the different levels of filtering (from 0.5 to 8 cycles/degree of visual angle) of the normally colored and luminance-only versions of color-diagnostic and color nondiagnostic scenes in Experiment 3. The figure reveals that colored scenes (the plain lines) elicited better categorization performance than the luminance-only version of the same scenes (the dashed lines). A further effect was observed in which the color-diagnostic colored scenes (the thick solid line) were systematically better recognized than the color-nondiagnostic scenes (the thin solid line) at very low spatial resolutions.

lines are the luminance-only representations of the same scenes. Figure 9 reveals a very different performance profile for these stimuli. At very low resolutions (below 1.5 cycles/degree) performance is almost identical. At higher resolutions, however, performance is better for color-nondiagnostic scenes, suggesting that these levels represent useful luminance cues, at least for the scenes considered.

The performance contrasts between colored and luminance stimuli at low resolution suggest that a spatial organization of coarse colored blobs is crucial information for early scene-recognition mechanisms. Objects cannot be identified at such resolution and so the configuration itself (*or* coarse spatial layout) appears to present sufficient classification information. However, a diagnostic organization of colored blobs cannot be the only factor that explains performance: color-nondiagnostic scenes were themselves better categorized than their luminance counterparts (see the difference between the thin lines, solid and dashed). Image segmentation could explain this discrepancy. With luminance as the only source, image segmentation depends on luminance contrasts. Colors add two supplementary sources of possible contrasts (the red–green and blue–yellow oppositions) and segmentation is better constrained. A better segmentation of blobs also promotes a more effective categorization of the scenes they represent.

In sum, Experiment 3 examined the nature of the chromatic and luminance information that can tap into scene-categorization mechanisms. It was found that the addition of colors to coarse luminance blobs enhanced their categorization. It was also shown that the contribution of colors across spatial scales depended on their diagnosticity. Thus, a coarse organization of diagnostically colored blobs (together with their easier segmentation) is in itself effective information for scene categorization.

GENERAL DISCUSSION

This paper reported three experiments that investigated the influence of color on real-world scene recognition. We reported converging evidence that colors play a primary role in speeded scene recognition and that coarse spatial layouts were powerful scene recognition cues.

Using a fast identification task, Experiment 1 revealed that color influences recognition when it is diagnostic of a category. The addition of normal colors to an achromatic stimulus facilitated its naming, whereas the addition of abnormal colors interfered with naming. In contrast, colors had no observable effects on the naming of color-nondiagnostic scenes. Turning to a verification task in Experiment 2, we also found selective effects of diagnostic colors. Subjects were faster to judge the category membership of color-diagnostic scenes when these were properly colored. In contrast, colors had no observable effects on the verification of color-nondiagnostic scenes. These results with a verification task replicated the findings of the naming task in Experiment 1 and demonstrated that color diagnosticity, not the experimental task itself, was the main factor explaining the effects of color on scene recognition. Furthermore, interference effects obtained with negative trials in Experiment 2 confirmed that memorized diagnostic scene colors intervene in recognition. Turning to the cues responsible for these effects in Experiment 3, we found that the addition of colors to luminance cues enhanced scene categorization but only at a coarse spatial scale, particularly when colors were diagnostic. We therefore conclude that a coarse organization of diagnostically colored blobs could effectively support the categorization of complex visual scenes.

Color and Scene Representations

The consistent effects of color diagnosticity reported in our experiments suggest that people do represent scene colors in memory. The role of color as a primary scene recognition cue should therefore be reexamined (remember that people could correctly identify colored scenes at very low spatial resolutions). The absence of an effect for color-nondiagnostic scenes categories does not mean that their colors are not represented in memory. The methods we have used (reaction times in a naming task in Experiment 1 and in a verification task in Experiment 2) are not powerful enough to diagnose the absence of a color encoding in memory. That is, colors could still be memorized, but have so little predictive value that they do not have an observable effect on naming and verification latencies. Gegenfurtner, Sharpe, and Wichmann (1995) asked subjects to learn colored and black-and-white versions of the same scenes to be later tested on their familiarity with these scenic pictures. It was found that subjects performed better with memorized colored images, suggesting that they memorized color information. It would be interesting to test whether color diagnosticity interacts with such judgments of scene familiarity. Color-diagnostic scenes could be color-transformed to be more typical of their category and therefore be judged more familiar, even when subjects did learn scene exemplars with colors different from those of the tested exemplars.

Coarse Scene Layouts for Scene Recognition

The role of the global scene context in the recognition of local objects is an important issue in recognition and eye movement studies (see Henderson & Hollingworth, 1999, for a thorough review). The notion of a *scene context*, *scene ‘‘gist,’’* or *scene spatial layout*, however, remains to be precisely defined (Rensink, 2000). In fact, it is only recently that its role in the recognition of complex pictures has been acknowledged (Epstein & Kanwisher, 1998; Kersten, 1997; Schyns & Oliva, 1994; Sanocki & Epstein, 1997). One goal of this study was to further specify the information content of the scene gist and to demonstrate its sufficiency for scene categorization. Remember that in Experiment 3, natural scenes (color-diagnostic and color-nondiagnostic) were both recognized from coarse colored information.

Sanocki and Epstein (1997) suggested that the spatial layout was a representation including information about the ground plane (its extent and location), the positions of component objects and surfaces, and the distance relations within the scene. It appears that organization of colored blobs on a coarse scale represents some of this information. The sky and the ground plane can be differentiated from the colored blobs, and the color contrasts between the blobs on the ground plane can be used to measure the positions of the main objects and surfaces and their relative relations within the image. It is clear from the results of all our experiments (but particularly from those of Experiment 3) that the addition of color facilitates the initial segmentation of the image.

Note that the initial recognition information forming the gist is specific to the scene itself. It is not derived from the identity of the objects (textural information, detailed contours, and cast shadows are not well represented on coarser scales). The role of coarse scene backgrounds in object identification clearly deserves further exploration because the identification of the former could facilitate the processing of the latter, both in humans and machines (e.g., Oliva, Torralba, Guerin-Dugue, & Herault, 1999; Swain, & Ballard, 1991; Vailaya, Jain, & Zhang, 1998).

CONCLUDING REMARKS

Our results with naturalistic scenes suggest that colors should not be dismissed in recognition studies. Colored blobs can mediate the recognition of scenes without prior recognition of their objects. Color is therefore an important property of the relevant spatial layout that mediates express scene recognition.

APPENDIX 1

Three variables can represent all colors of the visible spectrum. Three dimensions can therefore represent each colored pixel of a digitized image. In 1931, the *Comission Internationale de l'Eclairage* (CIE) set up a number of different standards for encoding colors. The best known is probably Red Green Blue (RGB), which encodes each pixel as a linear combination of three monochromatic dimensions. One problem with RGB is that it does not separate luminance from chromatic information: changing the colors of a pixel also modifies its luminance. To perform this separation, CIE proposed another basis, called *XYZ*, immediately derivable from RGB, in which *Y* represents luminance and *X* and *Z* are two additional axes from which color can be readily extracted. The following matrix multiplication summarizes the linear transformation between the *XYZ* and RGB spaces (see Wysczecki & Stiles, 1982):

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.412453 & 0.357580 & 0.180423 \\ 0.212671 & 0.715160 & 0.072169 \\ 0.019334 & 0.119193 & 0.950227 \end{bmatrix} \begin{bmatrix} \mathrm{R} \\ \mathrm{G} \\ \mathrm{B} \end{bmatrix}. \qquad (1)$$

If it is practical to separate luminance and chromatic information, it is psychologically interesting when this separation mirrors perception. In 1976, CIE proposed $L^*a^*b^*$, a 3D color basis derived from *XYZ*, which explicitly separates luminance on one dimension (L^*) from chroma on the two remaining dimensions (a^*b^*). Figure A1.1 summarizes the organization of the space for a fixed luminance value. The a^* axis represents the green/red opposition (green for $a^* < 0$ and $b^* > 0$ and red for $a^* > 0$ and $b^* = 0$), whereas

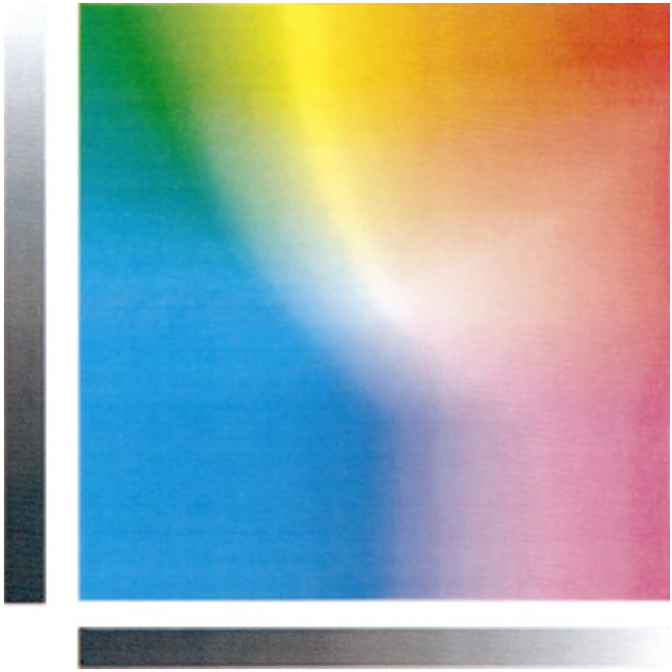


FIG. A1.1. The $L^*a^*b^*$ color space for a fixed luminance level. We used $L^*a^*b^*$ in our experiments to synthesize new stimuli and control the diagnosticity of scene colors.

b^* represents the blue/yellow opposition (blue for $a^* = 0$ and $b^* < 0$, yellow for $a^* = 0$ and $b^* > 0$). Stimuli for which $a^* = b^* = 0$ are achromatic.

The following nonlinear equations produce $L^*a^*b^*$ coordinates from XYZ.

For $(X/X_0, Y/Y_0, Z/Z_0) > 0.01$,

$$L^* = 116\left(\frac{Y}{Y_0}\right)^{1/3} - 16$$

$$a^* = 500\left[\left(\frac{X}{X_0}\right)^{1/3} - \left(\frac{Y}{Y_0}\right)^{1/3}\right] \tag{2}$$

$$b^* = 200\left[\left(\frac{Y}{Y_0}\right)^{1/3} - \left(\frac{Z}{Z_0}\right)^{1/3}\right],$$

where $[X_0 = 0.950456\ Y_0 = 1.0\ Z_0 = 1.088754]$ is the so-called ''CIE Illuminant D65,'' also commonly known as ''standard daylight white.''

APPENDIX 2

In all of our experiments, color-diagnostic categories were chosen with the constraint that their pixels projected into clearly distinguishable clusters in a^*b^*. Color-nondiagnostic categories were chosen so as not to form clear-cut clusters in a^*b^*. Thus, the overlap of the projections of categories in a^*b^* (their color histograms) can be used to control the diagnosticity of scene colors in the experiment. To compute the projections of categories in color space, we first divided a^*b^* in a lattice of 80 x 80 equally spaced bins. For each category, we retrieved the a^*b^* bin of each pixel of each exemplar by transforming the pixel from RGB to $L^*a^*b^*$, as detailed in Eqs. (1) and (2) of Appendix 1. We then normalized the outcome by summing the pixels in each bin and dividing the total per bin by the total number of pixels per category. This technique produced a different color histogram per category in a^*b^*. To measure the distinctiveness of the color histograms, we computed their pairwise correlations. Note that pairwise correlations assume an underlying Euclidian color space. Remember that $L^*a^*b^*$ is perceptually homogeneous for long distances in the space. Thus, our choice of scene categories that are distant from one another in $L^*a^*b^*$ licenses the use of pairwise correlations between color histograms to measure similarities of scene colors in this space. This computation requires that the 80×80 dimensional color histogram be transformed into a 80^2 normalized vector. It is important to stress what this metric really does: It compares the frequencies of each color, summed across the pixels and scenes of two categories without respect to the spatial location of the pixels of any given color.

Tables A2.1 and A2.2 provide the correlations for the color-diagnostic and color nondiagnostic categories of Experiment 1, respectively. Average pairwise correlations were .15 for color-diagnostic categories ($SD = .17$, see Table A2.1) and .8 ($SD = .05$) for color-nondiagnostic categories, indicating

TABLE A2.1
Correlations between the Color Histograms of the Color-Diagnostic Scenes in Experiment 1[a]

	Canyon	*Desert*	*Coast*
Forest	.06[b]	.05	.17
	.18	**.19**	**.14**
Canyon		.48	.04
		.39	**.07**
Desert			.12
			.2

[a] Numbers in bold refer to the correlations of the abnormally colored versions of the scenes [Mean = .19 (SD = .11)].

[b] Mean = .15 (SD = .17).

TABLE A2.2
Correlations between the Color Histograms of the Color-Nondiagnostic Scenes in Experiment 1[a]

	City	*Room*	*Road*
Shop	.71[b]	.76	.32
	.56	**.62**	**.63**
City		.85	.83
		.88	**.83**
Room			.8
			.86

[a] Numbers in bold refer to correlations of the abnormally colored versions of the scenes [Mean = .72 (*SD* = .140].
[b] Mean = .8 (*SD* = .05).

TABLE A2.3
Correlations between the Color Histograms of the Color-Diagnostic Scenes in Experiment 2[a]

	Canyon	*Desert*	*Coast*	*Field*	*Forest*	*Garden*	*Valley*
Beach	.18	.22	.42	.3	.39	.3	**.53**
Canyon		.38	**.01**	.07	.03	.14	.07
Desert			.05	**.16**	.02	.03	.09
Coast				.14	.18	.01	.36
Field					.39	.44	.44
Forest						**.46**	.6
Garden							.38

[a] Numbers in bold refer to scene pairs chosen to compose the contingency table of negative trials (see Table 1). [Mean = .24 (*SD* = .18)].

TABLE A2.4
Correlations between the Color Histograms of the Color-Nondiagnostic Scenes in Experiment 2[a]

	City	*Room*	*Road*	*Bath*	*Bedroom*	*Kitchen*	*Restaurant*
Shop	.59	.76	.37	.57	.51	.62	.68
City		.67	.43	.36	.35	.72	.51
Room			.48	.57	.61	.83	.71
Road				.44	.39	.47	.51
Bath					.66	.51	.62
Bedroom						.53	.69
Restaurant							.61

[a] Mean = .56 (*SD* = .13).

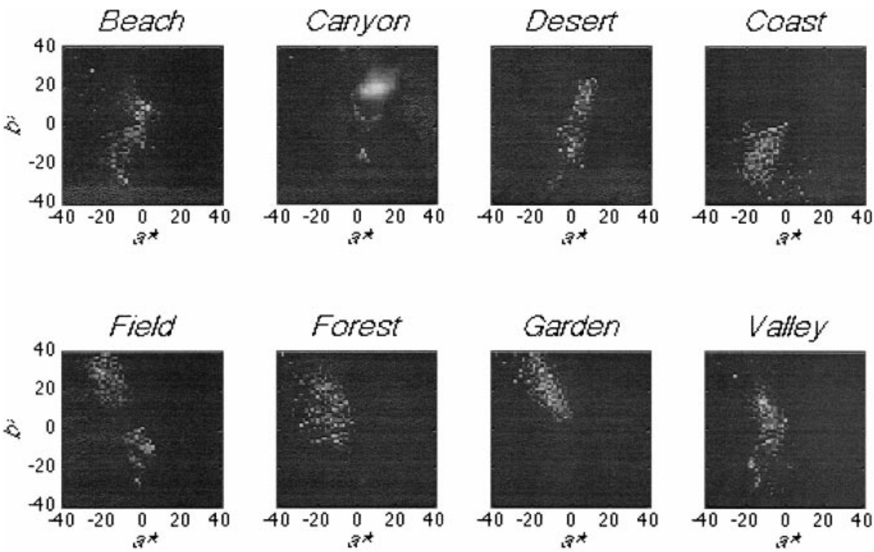


FIG. A2.1. The color histograms of the 8 color-diagnostic categories used in Experiment 2.

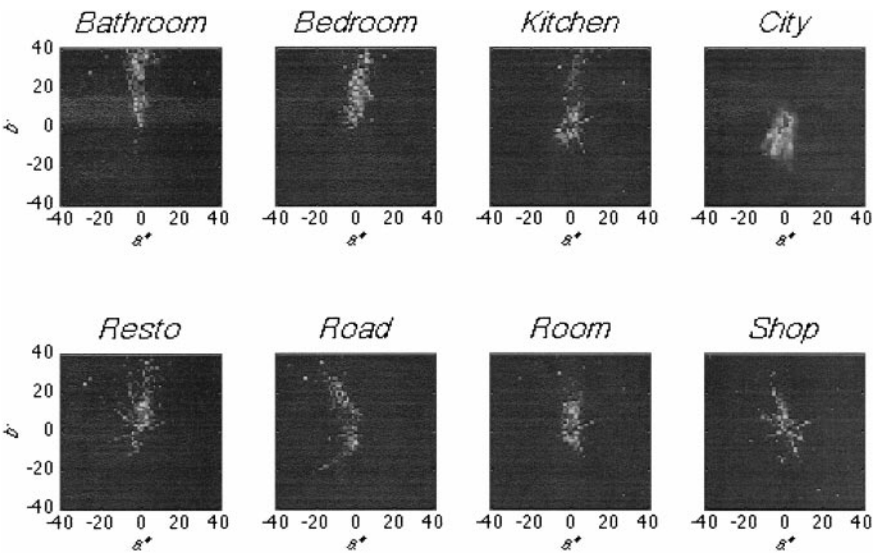


FIG. A2.2. The color histograms of the 8 color-nondiagnostic categories used in Experiment 2. A comparison with Fig. A2.1 reveals that the color-nondiagnostic categories have more overlap in their color histograms than color-diagnostic categories.

a low vs high overlap for color-diagnostic scenes vs -nondiagnostic categories, respectively. The bold numbers represent the pairwise correlations of color-transformed scenes. Their average was of .8 (SD = .05) for color-diagnostic scenes and .72 (SD = .14) for color-nondiagnostic scenes.

Tables A2.3 and A2.4 provide the correlations for the color-diagnostic and color-nondiagnostic categories of Experiment 2, respectively. Average pairwise correlations between color histograms were comparatively higher (mean = .56, SD = .13) for color-nondiagnostic than for color-diagnostic (mean = .24, SD = .18) categories, indicating a greater color overlap in the former. Figures A2.1 and A2.2 present the projections of the categories in a^*b^*.

APPENDIX 3

Experiment 1 was fully replicated with a design in all points identical except that stimuli were presented for 30 ms (instead of 120 ms). With 16 participants, the trends were in all points similar to those observed with 120-ms exposures. A two-way ANOVA on the categorization latencies with Nat vs Art categories and Lum vs Nat vs Abn stimuli also revealed a significant main effect category type [$F(1, 15) = 16.27, p < .001$], a significant effect of stimulus condition [$F(2, 30) = 27.68, p < .0001$], and a significant inter-

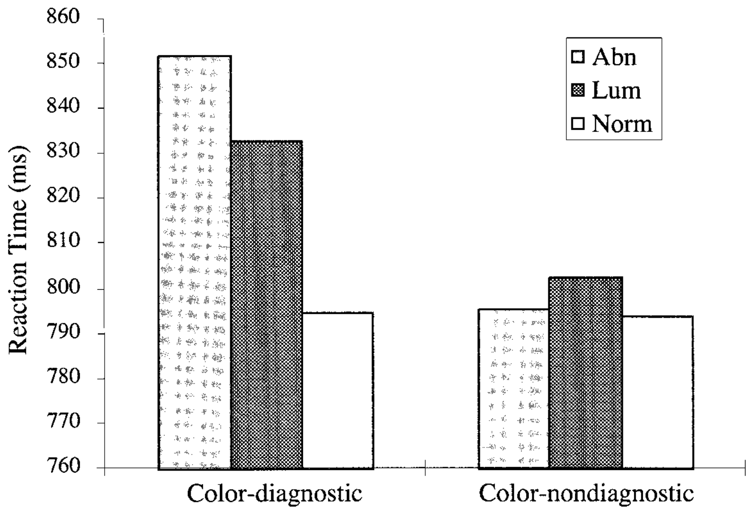


FIG. A3.1. Subjects' naming reaction times in the 30-ms replication of the within-subjects design of Experiment 1. Performance was very similar across luminance-only, normally colored, and abnormally colored versions of the color-nondiagnostic scenes. Facilitation of normally colored pictures and interference of abnormally colored pictures were observed for color-diagnostic scenes.

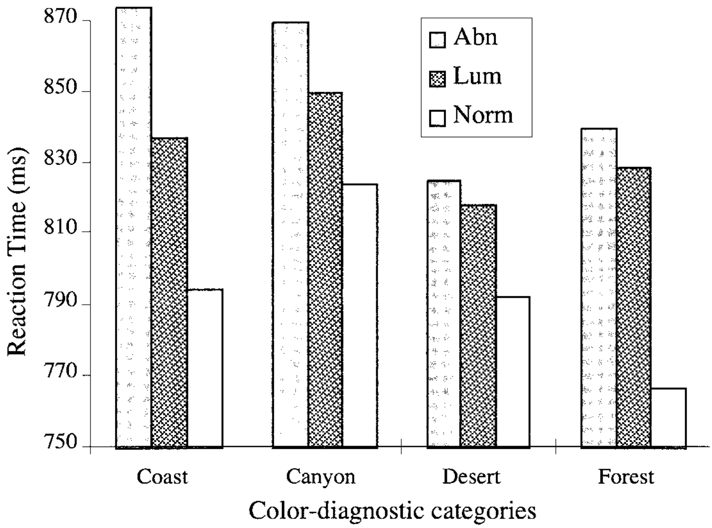


FIG. A3.2. The decomposition of the reaction times for each color-diagnostic category of the 30-ms replication of Experiment 1. The trends of facilitation of normal colors and interference of abnormal colors applied to each tested category.

action [$F(2, 30) = 25.69$, $p < .0001$; see Fig. A3.1]. As in experiment 1, the reported effects of color diagnosticity were true of all tested categories (see Fig. A3.2).

REFERENCES

Aginski, V., & Tarr, M. J. (2000). How are different properties of a scene encoded in visual memory? *Visual Cognition,* **7,** 147–162.

Biederman, I. (1981). On the semantics of a glance at a scene. In M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual organization* (pp. 213–233). Hillsdale, NJ: Erlbaum.

Biederman, I., & Ju, G. (1988). Surface versus edge-based determinants of visual recognition. *Cognitive Psychology,* **20,** 38–64.

Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: Detecting and judging objects undergoing relational violations. *Cognitive Psychology,* **14,** 143–177.

Billock, V. A. (1995). Cortical simple cells can extract achromatic information from the multiplexed chromatic and achromatic signals in the parvocellular pathways. *Vision Research,* **35,** 2359–2369.

Boyce, S. J., Pollatsek, A., & Rayner, K. (1989). Effect of background information on object identification. *Journal of Experimental Psychology: Human Perception and Performance,* **15,** 556–566.

Breitmeyer, B. G. (1984). *Visual masking: An integrative approach.* New York: Oxford Univ. Press.

Cavanagh, P. (1996). When colours move. *Nature,* **379,** 26.

Cavanagh, P., & Ramachandran, V. S. (1988). Structure from motion with equiluminous stimuli. *Annual Meeting CPA, Montreal*.

Connolly, C. (1996). The relationships between colour metrics and the appearance of three-dimensional coloured objects. *Color Research and Application,* **21,** 331–337.

Costen, N. P., Parker, D. M., & Craw, I. (1994). Spatial content and spatial quantisation effects in face recognition. *Perception,* **23,** 129–146.

Cropper, S. J., & Derrington, A. M. (1996). Rapid colour-specific detection of motion in human vision. *Nature,* **379,** 72–74.

Damasio, A., Yamada, T., Damasio, H., Corbett, J., & McKee, J. (1980). Central achromatopsia: behavioral, anatomic, and physiological aspects. *Neurology,* **30,** 1064–1071.

Delorme, A., Fabre-Thorpe, M., Richard, G., Fize, D., & Thorpe, S. J. (1998). Ultra-rapid visual categorisation of natural scenes is colour-blind. *Perception,* **27** (Suppl.), 170.

Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature,* **392,** 598–601.

Fiorentini, A., Maffei, L., & Sandini, G. (1983). The role of high spatial frequencies in face perception. *Perception,* **12,** 195–201.

Friedman, A. (1979). Framing pictures: The role of knowledge in automatized encoding and memory for gist. *Journal of Experimental Psychology: General,* **108,** 316–355.

Gegenfurtner, K. R., Sharpe, L. T., & Wichmann, F. A. (1995). The contribution of color to recognition memory in normal and color-deficient observers. *Perception,* **24** (Suppl.), 12.

Gegenfurtner, K. R. (1998). Sensory and cognitive contribution of color to the recognition of natural scenes. *IOVS,* **39,** s156.

Henderson, J. M. (1992). Object identification in context: The visual processing of natural scenes. *Canadian Journal of Psychology,* **46,** 319–341.

Henderson, J. M., & Hollingworth, A. (1999). High-level scene perception. *Annual Review of Psychology,* **50,** 243–271.

Hollingworth, A., & Henderson, J. M. (1998). Does consistent scene context facilitate object perception? *Journal of Experimental Psychology: General,* **127,** 398–415.

Intraub, H. (1997). The representation of visual scenes. *Trends in Cognitive Science,* **1,** 217–222.

Joseph, J. E., & Proffitt, D. R. (1996). Semantic versus perceptual influences of color in object recognition. *Journal of Experimental Psychology: Learning, Memory and Cognition,* **22,** 407–429.

Kersten, D. (1997). Perceptual categories for spatial layout. *Philosophical Transactions of the Royal Society of London, Series B,* **352,** 1155–1163.

Livingstone, M. D., & Hubel, D. H. (1987). Psychophysical evidence for separate channels for the perception of form, color, movement, and depth. *Journal of Neuroscience,* **7,** 3416–3468.

Logothetis, N. K., Schiller, P. H., Charles, E. R., & Hurlbert, A. C. (1989). Perceptual deficits and the role of color-opponent and broad-band channels in vision. *Science,* **247,** 214–217.

Marr, D. (1982). *Vision*. San Francisco: Freeman.

McIlhaga, W., Hine, T., Cole, G. R., & Snyder, A. W. (1990). Texture segregation with luminance and chromatic contrast. *Vision Research,* **30,** 489–498.

Mullen, K. T. (1985). Contrast sensitivity of human color vision to red–green and blue–yellow chromatic gratings. *Journal of Physiology (London),* **359,** 381–400.

Oliva, A., & Schyns, P. G. (1996). Color influences fast scene categorization. *Proceedings of the XVIII Meeting of the Cognitive Science Society* (pp. 239–242). Hilldsale, NJ: Erlbaum.

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.

Oliva, A., Torralba, A., Guerin-Dugue, A., & Herault, J. (1999). Global semantic classification using power spectrum templates. *The Challenge of Image Retrieval,* Proceedings of Electronic Series in Computer Science, New York: Springer-Verlag.

Ostergaard, A. L., & Davidoff, J. B. (1985). Some effects of color on naming and recognition of objects. *Journal of Experimental Psychology: Learning, Memory and Cognition,* **11,** 579–587.

Parker, D. M., Lishman, J. R., & Hughes, J. (1992). Temporal integration of spatially filtered visual images. *Perception,* **21,** 147–160.

Parker, D. M., Lishman, J. R., & Hughes, J. (1996). Role of coarse and fine information in face and object processing. *Journal of Experimental Psychology: Human Perception and Performance,* **22,** 1448–1466.

Parraga, C. A., Brelstaff, G., Troscianko, T., & Moorhead, I. (1998). Color and luminance information in natural scenes. *Journal of Optical Society of America A,* **15.** 3.

Potter, M. C. (1975). Meaning in visual search. *Science,* **187,** 965–966.

Potter, M. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory*, 2, 509-522.

Price, C. J., & Humphreys, G. W. (1989). The effects of surface detail on object categorization and naming. *The Quartely Journal of Experimental Psychology,* **41A,** 797–828.

Rensink, R. (2000). The dynamics representation of scenes. *Visual Cognition,* **7,** 17–42.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyce-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology,* **7,** 273–281.

Sanocki, T., & Epstein, W. (1997). Priming spatial layout of scenes. *Psychological Science,* **8,** 374–378.

Sanocki, T., Bowyer, K. W., Heath, M. D., & Sarkar, S. (1998). Are edges sufficient for object recognition? *Journal of Experimental Psychology: Human Perception and Performance,* **24,** 340–349.

Schyns, P. G., & Oliva, A. (1994). From blobs to boundary edges: Evidence for time and spatial scale dependant scene recognition. *Psychological Science,* **5,** 195–200.

Schyns, P. G., & Oliva, A. (1997). Flexible, diagnosticity-driven, rather than fixed, perceptually determined scale selection in scene and face recognition. *Perception,* **26,** 1027–1038.

Schyns, P. G., & Oliva, A. (1999). Dr. Angry and Mr. Smile: When categorization flexibly modify the perception of faces in rapid visual presentation. *Cognition,* **69,** 243–265.

Sergent, J. (1986). Microgenesis of face perception. In D. H. Ellis, M. A. Jeeves, F. Newcombe, & A. Young (Eds.), *Aspects of face processing* (pp. 17–73). Dordrecht: Martinus Nijhoff.

Swain, M., & Ballard, D. (1991). Color Indexing. *International Journal of Computer Vision,* **7,** 11–32.

Tanaka, J. W., & Presnell, L. M. (1999). Color diagnosticity in object recognition. *Perception & Psychophysics,* **61,** 1140–1153.

Tarr, M. J., & Pinker, S. (1989). Mental rotation and orientation-dependence in shape recognition. *Cognitive Psychology,* **21,** 233–282.

Vailaya, A., Jain, A., & Zhang, H. J. (1998). On image classification: city images vs. landscapes. *Pattern Recognition,* **31,** 1921–1936.

Wurm, L. H., Legge, G. E., Isenberg, L. M., & Luebker, A. (1993). Color improves object recognition in normal and low vision. *Journal of Experimental Psychology: Human Perception and Performance,* **19,** 899–911.

Wyszecki, G., & Stiles, W. S. (1982). *Color science: Concept and methods, quantitative data and formulae* (2nd ed.). New York: Wiley.

(Accepted August 1, 1999)