Journal of Experimental Psychology:
Human Perception and Performance
1993, Vol. 19, No. 4, 899–911

0096-1523/93/$3.00

Color Improves Object Recognition in Normal and Low Vision

Lee H. Wurm, Gordon E. Legge, Lisa M. Isenberg, and Andrew Luebker

Does color improve object recognition? If so, is the improvement greater for images with low spatial resolution in which there is less shape information? Do people with low visual acuity benefit more from color? Three experiments measured reaction time (RT) and accuracy for naming food objects displayed in 4 types of images: gray scale or color, and high or low spatial resolution (produced by blur). Normally sighted Ss had faster RTs with color, but the improvement was not significantly greater for images with low spatial resolution. Low vision subjects were also faster with color, but the difference did not depend significantly on acuity. In 2 additional experiments, it was found that the faster RTs for color stimuli were related to objects' prototypicality but not to their color diagnosticity. It was concluded that color does improve object recognition, and the mechanism is probably sensory rather than cognitive in origin.

What is the function of color vision? Does it facilitate object recognition? Despite the impressive neural resources devoted to color vision and the perceptual salience of color, people with color-vision deficits (about 4% of the population) experience very few difficulties. In fact, many colorblind people become aware of their deficit only when they take a color test (Steward & Cole, 1989).

It is likely that color vision evolved in response to selection pressures very different from those faced by modern humans. Polyak (1957) suggested that color vision coevolved with brightly colored plants. In a compelling photograph, he showed that ripe, red berries are easily seen against a background of green leaves, whereas the unripe green berries are much harder to see (Polyak, 1957, Figure 518-A). In Polyak's view, color vision evolved to facilitate food gathering, involving search and recognition of natural objects. Consistent with this view, vertebrates who live in different environments exhibit variation in both the anatomy of color vision and their behavioral response to color (Jacobs, 1981).

Color may play a role in scene segmentation. Walls (1942, p. 463) suggested that color promotes the perception of contour. Color differences, like luminance differences, can be used to segment images into regions containing information about individual objects. Contours defined by color differences may provide more reliable information about object shape in the natural world than luminance contours because shadows and occlusion boundaries also produce luminance contours (De Valois & Switkes, 1983; K. De Valois, personal communication, September 18, 1990).

Knowledge of the colors of familiar objects may help in recognition. Objects vary in the degree to which their colors are "diagnostic" (Biederman & Ju, 1988). Although virtually no object can be recognized on the basis of its color alone, the colors of some objects (e.g., an apple) are less arbitrary than others (e.g., a car). In an RT task, Biederman and Ju found no advantage for objects rated as having highly characteristic colors. We return to the question of color diagnosticity in the final paragraph of the introduction and in Experiment 5.

There is disagreement whether color facilitates object recognition. Markoff (1972) measured RTs for subjects to decide which of three targets (tank, jeep, or soldier) was present in a black-and-white or color slide. The targets were hidden in real-world backgrounds. He blurred the slides to evaluate the interaction of spatial resolution and color. He found that RTs were shorter (and error rates lower) for the color slides, and the advantage of color over black-and-white performance increased with greater blur. These results indicate that color is helpful in a search task and that color may be more helpful when shape information is degraded.

Ostergaard and Davidoff (1985) measured RTs for subjects to name fruits and vegetables shown in slides. Half of the slides were color, and half were black and white. Ostergaard and Davidoff found that the objects were named more quickly in the color slides.

Biederman and Ju (1988) measured RTs for subjects to name objects displayed as color photos or simple line drawings. Twenty-five of the 29 targets were man-made objects with well-specified geometrical forms, such as cameras and hairdryers. (The four exceptions were mushroom, apple, banana, and fish). There were no systematic differences in RTs for the color photos and line drawings. Biederman and Ju concluded that color does not play a direct role in object recognition except when reliable shape cues are absent or degraded, but color may play an indirect role through scene segmentation. According to their view,

---

Lee H. Wurm (now at the Department of Psychology, State University of New York at Stony Brook), Gordon E. Legge, Lisa M. Isenberg (now at the Department of Psychology, University of Wisconsin—River Falls), and Andrew Luebker, Department of Psychology, University of Minnesota.

This work was supported by National Institutes of Health Grants EY02934 and EY02857, U.S. Air Force Office of Scientific Research Grant 90–0274, and a grant from the University of Minnesota's Undergraduate Research Opportunities Program.

We thank Irving Biederman for useful discussion of many aspects of this article, Daniel Kersten for discussion of the information-theory analysis of color diagnosticity, Paul Beckmann for help in digitizing the stimuli, and Hiromi Akutsu and Yuanchao Gu for help in display calibration.

Correspondence concerning this article should be addressed to Gordon E. Legge, Department of Psychology, University of Minnesota, Minneapolis, Minnesota 55455.

if shape information is available, primary (fastest) access to an object's mental representation involves matching to a structural description on the basis of a few simple volumetric primitives (Biederman, 1987).

A problem with all of these studies is that luminance characteristics varied across the color conditions. In the Markoff (1972) and the Ostergaard and Davidoff (1985) studies, the distributions of luminance were not matched in the color and black-and-white slides. In the Biederman and Ju (1988) study, line drawings were compared with color photos. Visual analysis of the color photos may have been at a starting disadvantage because of greater difficulty in edge extraction compared with line drawings.

One interpretation of the foregoing studies is that color improves object recognition more when spatial resolution is low (blur) or when shape information is less specific (fruits and vegetables vs. man-made objects). One major purpose of our research was to examine the hypothesis that color and shape information interact in object recognition; that is, color facilitates object recognition more when spatial resolution is low. Psychophysical and computational studies show that chromatic contrast sensitivity is confined to a lower spatial-frequency range than luminance contrast sensitivity (Derrico & Buchsbaum, 1991; Kelly, 1983; Mullen, 1985). These studies are consistent with the hypothesized interaction between color and blur, assuming that low-frequency chromatic contrast sensitivity can be recruited to improve object recognition when high-frequency information is removed by blur.

Our interest in the interaction of color and spatial resolution stems from research in our laboratory on low vision. *Low vision* is any chronic visual disability, not correctable by glasses or contact lenses, that impairs everyday activities. Recent estimates indicate that there are more than three million Americans with low vision (Tielsch, Sommer, Witt, Katz, & Royall, 1990). Snellen acuity, corresponding to spatial resolution, is one way of classifying low vision subjects. People with best corrected Snellen acuities of about 20/60 have only mild deficits in spatial resolution, but people with acuities as low as 20/4000 can use coarse shape information. The lower the acuity, the less visible is texture and subtle shape information. A second purpose of our research was to examine the hypothesis that color improves object recognition more for people with very low acuity.

Confirmation of this hypothesis would point to the practical significance of color as a way of coding information for low vision. Kaiser (1972) made a similar point, and Sicurella (1977) described some of the practical uses of color in low vision. Knowlton and Woo (1989) pointed out that color is widely used in primary education for illustrating new concepts, so an understanding of the salience of color in low vision has educational implications as well.

Although little is known about the role of color in low vision object recognition, the effects of color on normal and low vision reading are fairly well understood (Knoblauch, Arditi, & Szlyk, 1991; Legge, Parish, Luebker, & Wurm, 1990; Legge & Rubin, 1986). The results of these studies indicate that color has no practical advantages for low vision reading. Their research with normally sighted subjects showed that rapid reading can occur with equiluminant text in which the letters are rendered by chromatic contrast only (e.g., green letters on a red background). This finding indicates that color can be used for pattern recognition in a high-speed information-processing task. As part of their study, Legge et al. (1990) investigated additivity of color contrast and luminance contrast in reading. They asked whether differences in chromaticity between letters and background would combine with luminance differences in enhancing performance. They found no evidence for interaction and inferred that subjects relied on signals in either a luminance or color pathway. This lack of interaction reinforces psychophysical findings of relatively independent processing of color and luminance (Cole, Stromeyer, & Kronauer, 1990) and may be related to findings of independent processing of shape and color (Isenberg, Nissen, & Marchak, 1990).

In our two main experiments (Experiments 2 and 3), normal and low vision subjects named images that appeared on a color monitor. The images were digitized from real objects. They were presented either in full color or as gray-scale images matched pixel by pixel in luminance with the color images. A hypothetical subject having no sensitivity to chromaticity differences but normal luminance spectral sensitivity would see no difference between the two sets of images.

We chose food items for our objects. Real-world objects vary in the degree to which they can be given structural (i.e., geometric) descriptions. Man-made objects like forks, chairs, and hammers have stereotypical shapes. Natural objects like berries, potatoes, branches, rabbits, or stones are less regular in shape but retain some structure. Sugar, sand, or water must be recognized through cues other than shape. Similarly, objects vary in their color specificity. We chose to use food objects because they have a wide range of colors and are broadly representative in shape and color of natural objects. It is likely that color vision evolved in response to functional interaction with natural objects, perhaps including food, so this class of objects may provide a favorable domain for revealing a role for color.

When the results of the two main experiments showed that color does speed up object recognition, we conducted two additional experiments to examine the origins of this effect. A category label like APPLE refers not only to a prototypical image with a given color and shape (red and round with a stem) seen from a preferred viewpoint but also to many alternative images. The APPLE category includes other colors (yellow, green) and various shape deformations (sliced or bitten). We refer to such image variations as *different poses*. In our study, we used 21 food categories with four poses each. Biederman and Ju (1988) argued that structural (shape) descriptions are primary mental representations and that color and other surface characteristics play a role only when shape information is degraded or uninformative. For natural objects like apples that can still be recognized under several transformations, it may be that object recognition will be fastest, and there will be least facilitation by color, for the most prototypical exemplars. In Experiment 4, subjects ranked the four poses of each food category for prototypicality. We used these rankings in conjunction with the RT data to determine whether color speeds up object recognition more for poses that are nonprototypical. With a similar method, Palmer, Rosch, and Chase (1981) obtained the most prototypical view (termed *canonical perspective*) from among a set of views for each of several common objects. They found that RTs were fastest for the most prototypical views.

Earlier, we distinguished between scene segmentation and diagnosticity as ways in which color might facilitate object recognition. Assuming scene segmentation is a bottom-up process, it should not depend on our knowledge of the colors of things. Diagnosticity, on the other hand, relies on memory. It might improve object recognition by restricting the set of possible alternatives. From an information-theory perspective, two factors determine whether color is diagnostic of an object: The color must be symptomatic of the object (e.g., "green" is a "symptom" of spinach), and not many other objects in the allowable domain should have the same color. Although green is certainly a reliable symptom of spinach, there are many other fruits and vegetables that are green; green is not "diagnostic" of spinach. In our final experiment (Experiment 5), we used subjective probability ratings to compute an information-theoretic measure of color diagnosticity for each of our food categories. We asked whether color speeds up recognition more for food categories with high color diagnosticity.

Experiment 1: Preliminary

Method

Subjects. There were 24 subjects. All had normal or corrected-to-normal visual acuity according to the Light House Distance Visual Acuity chart and normal color vision according to the Farnsworth-Munsell 100 hue test. They received either extra credit in a psychology course or cash for their time. All subjects were native English speakers.

Stimuli and apparatus. The stimuli were produced on a Conrac 7241 color monitor. The red, green, and blue guns of the monitor were connected to three Imaging Technology (Bedford, MA) FG-100 frame buffers in an IBM PC-AT computer. Each frame buffer had a resolution of 512 × 480 pixels and 256 gray levels. The Commission Internationale l'Eclairage chromaticity coordinates of the three phosphors, measured with a Minolta CS-100 Chroma Meter, were (.617, .352) for red, (.281, .602) for green, and (.146, .053) for blue. Each image consisted of 128 × 128 pixels (7.0 × 7.6 cm) displayed at the center of the screen. Pixel luminances ranged from 0.01 to 100 cd/m^2.

The stimuli were 100 cathode-ray tube (CRT) images of common food items. There were 21 food categories (Table 1), each having at least four different poses. The poses consisted of individual or group pictures (e.g., a single banana or a bunch) and common deformations of the item (e.g., a peeled banana and a sliced banana).

Images were digitized from real objects with a Sony color video camera, Model DXC-M3A. The backdrop was a piece of neutral gray paper, and the lighting was provided by two 3200-K bulbs in parabolic reflectors. The resulting images contained some shadows, and some of the more shiny objects had highlights. The shadows and highlights appeared natural. The video camera's iris setting was kept constant to equalize the luminance of the background in the images.

The images were digitized so as to nearly fill the 128 × 128 pixel region regardless of the physical size of the food items, so scale information was not available to the subjects. For each image, the red, green, and blue signals from the video camera were stored as three separate files of 8-bit bytes. For display, these three files were loaded into the three frame buffers and simultaneously displayed by the red, green, and blue guns of the color monitor. The result is a full-color image.

Video digitization followed by CRT display does not yield a perfect match in chromaticity coordinates between stimulus objects and displayed images. We evaluated these mismatches by comparing them with shifts in chromaticity that occur in the everyday world under different illuminants. For a selected point on each of 12 food objects, we measured the chromaticity coordinates of the full-color video image and of the actual objects for three sources of illumination: 3200-K bulbs, Philips F40CW 40-W overhead fluorescent room lighting, and natural sunlight. Examples of these measurements are shown in Figure 1, and the results are listed in Table 2. The scatter is substantial, but the chromaticity coordinates of the video images usually lie within the cluster. We conclude that the errors in chromaticity introduced by digitization produce changes within the everyday range.

One hundred gray-scale (achromatic) images were derived from the corresponding full-color images and were matched in luminance on a pixel-by-pixel basis. For each pixel, the individual luminance contributions of each red, green, and blue digitized gray level were predicted with "gamma" tables describing the gray-level-to-luminance relation for each of the three color phosphors. The appropriate red, green, and blue gray levels for the equivalent achromatic pixel were then computed by first fixing the red:green:blue ratio to produce a constant chromaticity of $x = .28$, $y = .29$, and then setting the sum of the red, green, and blue luminance contributions to match the luminance of the corresponding pixel in the color image.[1] The results of this procedure were checked photometrically with test patches and experimental stimuli to verify that corresponding regions in full-color and gray-scale images were matched for luminance. Table 3 shows examples of these measurements and indicates that the error was usually less than 10%.

Blurred versions of the full-color and gray-scale images were produced by placing a sheet of ground glass (25.5 cm wide × 13 cm high) 6 mm from the screen of the monitor.[2] (The surface of the screen was effectively flat across the extent of the 7 × 7.6-cm images.) Ground glass blurring acts as a low-pass filter that attenuates high spatial frequencies more than low spatial frequencies. It is radially symmetric in the sense that all orientations are affected equally (for more details, see Legge, Pelli, Rubin, & Schleske, 1985). The modulation transfer function (MTF) of the blur screen was estimated psychophysically by obtaining contrast thresholds for square-wave gratings with and without the blur screen. The cutoff of the MTF was approximately 6 cycles/degree. The blur screen produced shifts in chromaticity that were smaller than those caused by the different illuminants discussed in connection with Figure 1 and Table 2.

A uniform gray patch of the same size as the food images was shown on the screen whenever no food image was displayed. Its luminance was 7.94 cd/m^2, equal to the average luminance of all pixels of all the food images. Its presence ensured that the subject's level of light adaptation was appropriate for the stimulus set. Despite the relatively low luminance levels, the images appeared realistic, with good color.

---

[1] Chromaticity coordinates of $x = .28$ and $y = .29$ resulted when the maximum gray level was used for the red, green, and blue components. Subjectively, it appeared white.

[2] We began by using digital filtering of the images but abandoned this method because of undesirable contouring effects that resulted from gray-level quantization in dark regions of the images.

Table 1
Stimulus Items

Category	Poses
APPLE	Single red[a]; red group; single yellow; yellow group[b]; several red & yellow[b]; sliced red
BANANA	Single[a]; bunch[a]; single peeled; banana slices
BREAD	Slice of white; two slices of white stacked; ½ loaf white[a]; slice French; ½ loaf French[b]
CANTALOUPE	Whole; half[a]; slice (¼); group of slices
CARROT	Single[a]; bunch; slices; diced; sticks[b]
CELERY	Single stalk[a]; group of stalks; celery slice; many slices
COOKIE	Molasses[b]; chocolate chip[a]; several chocolate chip; sugar; several sugar
GRAPE	Green[b]; bunch of green[a]; red[b]; bunch of red; one red & one green; bunch of red and green
LEMON	Single[a]; group; both pieces of a halved; lemon slice
OLIVE	Green[b]; group of green[a]; black[b]; group of black; one black and one green[a]; group of black and green
ORANGE	Single[a]; group; both pieces of a halved; slice; group of slices[b]
PEA	Single[b]; pea pod[a]; group of pods; group of peas; group of peas and pea pods
PEANUT	Single in shell[a]; group in shells; one shelled; group shelled
PEAR	Single[a]; group; single slice; group of slices
PEPPER	Single green[a]; group of green; single red; one green and one red[b]
PICKLE	Whole[a]; group[b]; slice[b]; group of slices; pickle spear[a]; group of spears
PINEAPPLE	Whole[a]; chunk; pineapple ring; group of pineapple chunks
POTATO	Single spud[a]; group of spuds; potato chip; several potato chips
PRETZEL	Single[a]; group[b]; pretzel stick; group of sticks; group of whole and sticks[b]
TOMATO	Single red[a]; group; tomato slice[b]; group of tomato slices
WATERMELON	¼ melon; melon slice; both sides of a halved slice; ½ melon slice[a]

[a] Stimuli judged to be most prototypical in each category. There were ties in the categories banana, olive, and pickle. [b] Stimuli used in practice trials only.

Procedure. Experiment 1 was a preliminary experiment in which we measured the accuracy with which normally sighted subjects named the target stimuli. The RT was not recorded, and there was no time pressure. There were 6 subjects in each of four groups. Each subject saw all 100 images (in a random sequence) in one of the four conditions: full color and no blur, full color with blur, gray scale with no blur, and gray scale with blur.

Prior to testing, subjects read a list of the 21 food categories and were told that their task was to name the type of object (e.g., APPLE rather than GRANNY SMITH or FRUIT). They were told to take as long as they wished and to strive for accuracy. They were told that the objects would fill the viewing window on the screen (i.e., no scale information) and that items could appear alone, in groups, or in parts (e.g., a single banana, bunch of bananas, or sliced banana). Subjects were encouraged to guess if they were not sure. After each trial, subjects were told the correct answer.

Subjects were tested 1 at a time. They were seated in front of the Conrac monitor at a distance of 57 cm. The images subtended 7° × 7.6° of visual angle.

Any image missed by more than 2 subjects in the full-color unblurred condition of this preliminary experiment was designated for use as a practice image in the subsequent experiments. There were

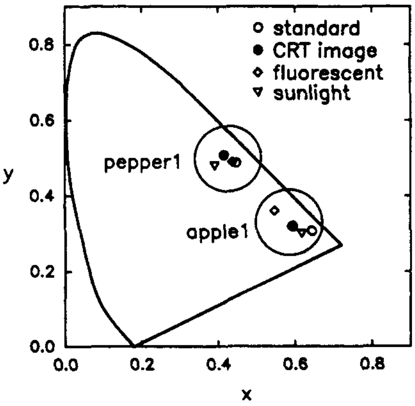

Figure 1. X and Y chromaticity coordinates of selected stimulus items under varying illuminants, represented in Commission Internationale l'Eclairage 1931 color space. ("Standard" illumination refers to the 3200-K bulbs, and "fluorescent" refers to Philips F40CW fluorescent lighting. See Table 1. CRT denotes cathode-ray tube.)

Table 2
Chromaticity Coordinates of Selected Food Items Under Varying Illuminants

	CRT image		3200–K		Fluorescent		Sunlight	
Food item	*x*	*y*	*x*	*y*	*x*	*y*	*x*	*y*
Green pepper	.415	.508	.447	.488	.437	.490	.391	.480
Potato	.429	.337	.511	.401	.477	.425	.438	.390
Red apple	.594	.320	.644	.308	.547	.361	.617	.301
Yellow apple	.389	.441	.507	.443	.490	.469	.453	.451
Red pepper	.628	.348	.667	.302	.559	.362	.611	.304
Lemon	.414	.443	.514	.437	.503	.463	.483	.451
Orange	.600	.369	.597	.385	.541	.437	.560	.411
Watermelon	.401	.276	.541	.367				
Banana	.436	.433	.523	.432				
Cantaloupe	.511	.389	.566	.401				
Red grape	.540	.350	.533	.361				
Carrot	.568	.360	.576	.378				

Note. CRT denotes cathode-ray tube.

16 such images, leaving 84 good images for the RT measurements. Error rates, given below, refer to performance on the 84 good images only.

Results and Discussion

Figure 2 shows the proportion of errors in the four stimulus conditions. Each point is the mean score of 6 normal subjects. Mean error rates ranged from 3% in the full-color unblurred condition to 31% for the gray-scale blurred condition. Following an arcsine transformation of the proportion of errors, an analysis of variance (ANOVA) revealed significant main effects of color, $F(1, 20) = 27.298, p < .001$, and blur, $F(1, 20) = 110.628$, $p < .001$, but no significant interaction, $F(1, 20) = 0.942, p > .25$.

Table 3
Comparing Luminances (cd/m^2) of Color and Gray-Scale Images

Image	Color	Gray scale
Corner of gray background (shadow)	1.69	1.50
Corner of gray background (direct lighting)	7.00	6.83
Reflection of apple 1[a]	42.7	42.6
Shadow of apple 1	0.59	0.40
White portion of banana 2	72.9	71.6
Banana peel (banana 2)	60.5	60.2
Blue test pattern[b]	4.69	5.07
Green test pattern	34.5	34.1
Red test pattern	10.3	11.4
Yellow test pattern	70.0	70.8
Magenta test pattern	25.9	26.0
Cyan test pattern	73.3	74.4

[a] Numbers refer to particular images (poses) within a food category. [b] Test patterns were rectangles displayed with various combinations of the monitor's red, green, and blue guns.

Experiment 2: RTs

Method

Subjects. Forty-eight different subjects participated in Experiment 2. Other subject characteristics are like those in Experiment 1.

Stimuli and apparatus. The same stimuli described in Experiment 1 were used. The RTs to name the images were measured with a voice-keyed switch. The internal 1.19-MHz clock timer was programmed to provide more accurate timing than the default timing mode used by the computer operating system.[3]

Procedure. The instructions were similar to those in Experiment 1 except that subjects were told to name the images as quickly and accurately as possible. After a warning tone, the experimenter initiated the image presentation, which lasted until the voice key detected the response. After each trial, the correct response and RT were displayed.

The 84 good images from Experiment 1 were arranged in a single, pseudorandom order, subject to the constraint that each of the 21 food categories appear at least once before the halfway point of the image list. Color and resolution were within-subjects variables; that is, each subject was exposed to all four color-by-resolution conditions, with 21 trials in each of four blocks. A pair of subjects was assigned to each of the 24 possible orderings of the four conditions, giving 48 subjects. One member of each pair viewed the images in forward numerical order, and the other viewed the images in reverse order. Four practice images were shown prior to each of the four blocks.

Results and Discussion

Figure 3 shows RTs in the four conditions. Each point is the average of 48 subjects' median values. Latencies were

---

[3] At the onset of each trial, the computer waited for the start of a new video frame from the video frame buffer. At this point, the software read the timer and immediately displayed the image. When the subject made a verbal response, the signal from the microphone activated a comparator switch. After detecting the signal on a serial RS-232 status line, the software again read the timer to compute the RT interval.

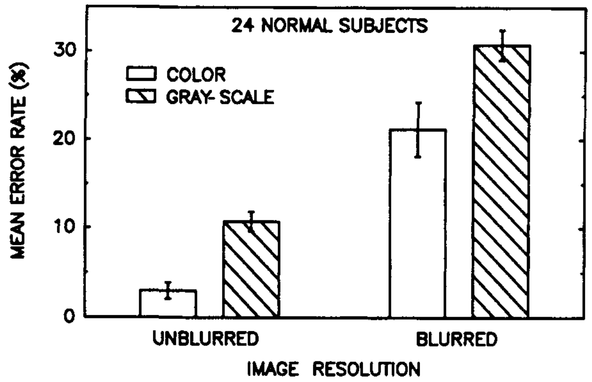

Figure 2. Percentage of images incorrectly identified for the four stimulus conditions in Experiment 1, with standard error bars. (Reaction times were not measured in this experiment.)

shortest in the full-color unblurred condition (795 ms) and longest in the gray-scale blurred condition (1068 ms). An ANOVA was performed on the log median RTs.[4] There were significant main effects of color, $F(1,47) = 34.866, p < .001$, and blur, $F(1, 47) = 47.7, p < .001$, but no significant interaction, $F(1, 47) = 1.337, p > .25$.

Percentage error rates are also shown in Figure 3. They ranged from 6% in the full-color unblurred condition to 37% in the gray-scale blurred condition and increased with increasing RTs. These error rates were 2%–7% higher than those in Experiment 1, but in other respects the pattern was the same. Following an arcsine transformation, an ANOVA revealed significant main effects of color, $F(1, 47) = 55.171, p < .001$, and blur, $F(1, 47) = 117.8$, $p < .001$, but no significant interaction, $F(1, 47) = 1.423$, $p > .2$.

Lack of evidence for a significant interaction should be qualified by the fact that the statistical test was weak. From a power analysis, we estimated the powers of the interaction tests on RTs and transformed errors to be about 30%.

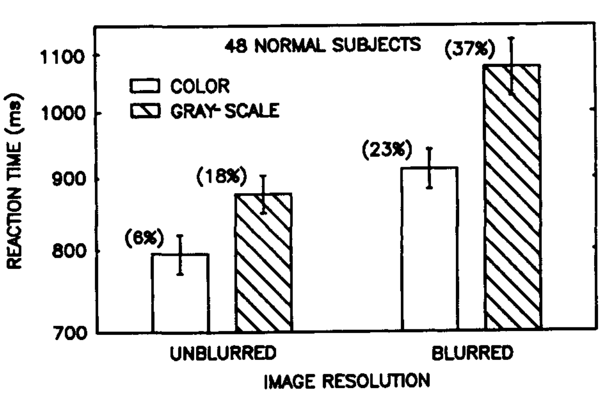

Figure 3. Mean of 48 subjects' median reaction times for the four stimulus conditions in Experiment 2, with standard error bars. (Reaction times are calculated for correct responses only. In parentheses are the mean error rates associated with each stimulus condition.)

There are two main findings from Experiment 1 and Experiment 2. First, color does improve recognition of food objects, whether measured as accuracy or RT. Second, we did not find the hypothesized interaction between color and spatial resolution; color appears to improve object recognition about equally for blurred and unblurred targets.

Experiment 3: RTs in Low Vision

Method

Subjects. Characteristics of our 16 low vision subjects are given in Table 4. Central-field status and ocular-media status are included because these are variables known to play a role in low vision reading. In particular, people with central scotomas usually read more slowly than acuity-matched subjects with intact central vision (Legge, Ross, Isenberg, & LaMay, 1992; Legge, Rubin, Pelli, & Schleske, 1985). We were interested in determining whether a similar difference exists for object recognition. Low vision subjects were selected from our roster of research participants on the basis of availability and the following loose criteria: (a) relatively normal color vision (D-15 test), (b) about an equal number of subjects with central-field loss and intact central vision, and (c) a similar distribution of acuities in the groups of subjects with intact central vision and central loss. Acuity tests and color tests were administered in our laboratory. Diagnoses and other medical information were obtained from subjects' ophthalmologists or optometrists.

Stimuli and apparatus. Stimuli and RT measurements were identical to those in Experiment 2, except that the ground glass was not used.

Procedure. The procedures were like those for Experiment 2 with the following exceptions. Image size was zoomed by a factor of two to 14° × 15.2°.[5] Because there was no ground-glass blurring, the 84 images were divided into two blocks of 42, one with full color and the other gray scale. Eight practice images were shown prior to each block. The 16 subjects were assigned to four groups of 4 subjects each on the basis of high or low acuity (on the basis of the median of the low vision sample) and on the presence or absence of central-field loss. The 4 members of each group each ran through the experiment in one of the following orders: Color–No Color × Forward–Backward list arrangement.

As in the previous experiments, there were no formal fixation instructions. Subjects with central scotomas were free to use off-foveal viewing of the stimuli.

Results and Discussion

We tested low vision subjects with unblurred images only. Figure 4 shows median RT as a function of acuity. Values for

---

[4] Variances of the median RTs in the four conditions were significantly different. After a log transformation, the variances were homogeneous.

[5] The enlarged images meant that low vision subjects covering a wider range of acuities could participate usefully in the study. Our purpose in studying low vision subjects was not to make a direct comparison with subjects with normal vision under identical viewing conditions but to treat low vision as a naturally occurring variation of spatial resolution.

Table 4
Characteristics of Low-Vision Subjects

Subject, age	Snellen acuity	Central field	Ocular media	Diagnosis	D-15 color test
A, 37	20/200	Intact	Clear	Detached retina, macular pucker	Normal
B, 50	20/250	Loss	Clear	Retinopathic lesioning	Normal
C, 74	20/250	Loss	Clear	Macular degeneration	Normal
D, 35	20/500	Loss	Clear	Retrolentalfibroplasia	Normal
E, 27	20/63	Intact	Clear	Congenital glaucoma	Normal
F, 40	20/160	Intact	Clear	Optic nerve hypoplasia	Normal
G, 42	20/800	Intact	Cloudy	Secondary corneal opacification	Normal
H, 43	20/200	Intact	Cloudy	Congenital cataracts, severe nystagmus	Normal
I, 30	20/400	Loss	Clear	Macular degeneration	Normal
J, 45	20/200	Loss	Clear	Atypical macular degeneration	Normal
K, 39	20/160	Loss	Clear	Macular degeneration	Normal
L, 31	20/50	Intact	Cloudy	Diabetic retinopathy, subretinal hemorrhages	Normal
M, 43	20/320	Intact	Cloudy	Detached retina, cataract	Failed, no axis
N, 36	20/500	Loss	Clear	Macular degeneration	Normal (1 minor error)
O, 35	20/100	Intact	Cloudy	Congenital cataracts, astigmatism, surgical aphakia	Normal
P, 81	20/250	Loss	Clear	Macular degeneration	Normal (1 minor error)

full-color images are indicated by the open symbols at the ends of the vertical stems, and values for the gray-scale images are indicated by the filled symbols. A two-way ANOVA on the log median RTs revealed significant main effects of color, $F(1, 14) = 18.093, p < .001$, and acuity, $F(1, 14) = 8.211, p < .02$. There was no significant interaction between color and acuity ($p = .199$). Our second hypothesis was that subjects with poorer acuity would benefit more from color in object recognition. This hypothesis was not confirmed.

The subjects with low vision had longer RTs than normal subjects and showed a greater advantage for color.[6] Mean low vision RT for colored images was 1531 ms (compared with 795 ms for normals), and for gray-scale images it was 1884 ms (compared with 876 ms for normals). Low vision subjects also had lower error rates for colored images ($M = 32.9\%$) than for gray-scale images ($M = 41.4\%$). The 353-ms color advantage for low vision subjects, together with lower error rates, showed a substantial functional advantage of color.

The low vision RTs were quite variable, probably because of the heterogeneity of visual disorders. A multiple-regression analysis on log RTs (mean of each subject's two medians), collapsed across color conditions, indicated that log decimal acuity accounted for 42.7% of the variance, $F(1, 14) = 10.44, p < .01$, with status of the central fields accounting for an additional 15.25% of the variance, $F(2, 13) = 8.956, p < .005$.[7, 8]

We conclude that (a) color speeds up object recognition for people with low vision (at least those with fairly normal color vision), (b) poor acuity slows down object recognition, and (c) acuity and color act independently in low vision object recognition.

Experiment 4: Prototype Analysis

Method

Subjects. There were 12 normally sighted subjects, none of whom participated in other experiments in this study. Other subject characteristics are like those in Experiment 1.

Procedure. The 84 images from the previous experiments were used again. In each of 21 trials, four full-color unblurred images from one food category appeared on the monitor in a 2 × 2 arrangement.[9] Subjects rank ordered each picture from 1 (*best example*) to 4 (*worst example*), writing values in a corresponding 2 × 2 grid of boxes on a score sheet. Neither ties nor blanks were allowed. Instructions were adapted from those used by Rosch,

---

[6] This comparison should be qualified by the fact that the low vision subjects were presented with larger images than those seen by the normal subjects.

[7] Central-field status (intact or lost) was treated as a binary indicator variable. A subject was classified as a *central loss* if there was an absolute scotoma (absence of pattern vision) covering all or part of the central 5° (diameter) of the visual field. Otherwise the subject was classified as *central intact.*

[8] Subject G with intact central vision had extremely long RTs With his data included, the mean latency of the central-loss group was 95 ms longer than the central-intact group, but his exclusion did not affect the outcome of significance tests reported here.

[9] We expect that the prototype scores would have been very similar to those we obtained if we had used gray-scale rather than full-color images. Consider the APPLE category, in which color might be expected to play a role in selecting a prototype. The single red apple was ranked more prototypical than the single yellow apple. But even in the corresponding gray-scale images, the darker skin on the red apple would likely result in its choice as most prototypical.

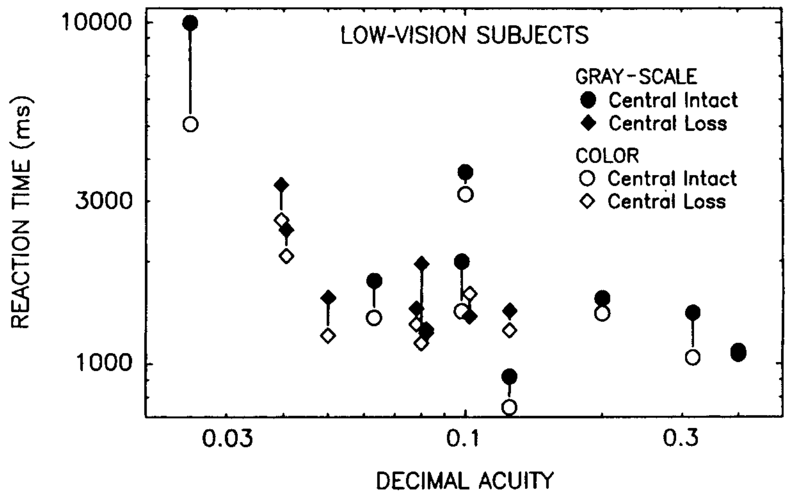

Figure 4. Median reaction times as a function of log decimal acuity for 16 low vision subjects in Experiment 3. (Decimal acuity is equal to the Snellen ratio [i.e., Snellen acuity of 20/20 equals decimal acuity of 1.0.] For each subject, the open symbol at the end of the vertical stem shows the reaction time for the full-color images, and the filled symbol shows the value for the gray-scale images. No blur conditions were used in this experiment.)

Mervis, Gray, Johnson, and Boyes-Braem (1976). The entire session lasted about 20 min.

For counterbalancing, subjects fell into four groups of 3 people. Each group was assigned a different random trial order. Image placement in the 2 × 2 grid in each trial was counterbalanced by rotating the grid one slot clockwise for each group.

Results and Discussion

Median prototype rankings were computed for each image. The image judged most prototypical in each of the 21 categories is tagged with an "a" in Table 1. For 14 of the 21 categories, 7 or more of the subjects agreed on the most prototypical image. In the remaining seven categories, there was no clear prototype. Median prototype rankings for these seven categories ranged from 1.5 to 2. Clear prototypes tended to be pictures of single, whole foods (for example, a single red apple) rather than bunches of items or piles of peeled, sliced, or chopped items.

Images ranked as more prototypical were recognized faster than those ranked as less prototypical. The correlation between median prototype score and mean RT for the complete set of 84 images was .481 ($p < .0001$). The link between perceptual prototype and RT has been documented by Palmer et al. (1981).

Prototypical images may be recognized faster because they contain characteristic shape information that subjects store in memory. Less prototypical images may have poorer shape cues, forcing greater reliance on color or other surface attributes. If so, we would expect the advantage of color to increase for less prototypical images. The correlation between normalized color advantage, which is equal to gray-scale RT minus full-color RT, all divided by the average of gray-scale and full-color RT, and median prototype score for the 84 images was .216, $p < .05$.[10] The weak but significant correlation means that the normalized color advantage was greater for items judged to be less prototypical.

We conclude that subjects are faster at recognizing images judged to be highly prototypical. But less prototypical images benefit more from color, that is, show a greater reduction in RT. These findings are consistent with the following view (Biederman & Ju, 1988): Prototypes are faster to recognize because they have more distinctive shapes. Nonprototypes are slower to recognize but benefit more from color because their recognition is more dependent on surface characteristics.

Experiment 5: Color Diagnosticity

Method

Subjects. There were 9 normally sighted subjects, none of whom participated in our other experiments. Other subject characteristics are like those in Experiment 1.

Procedure. We used an information-theory analysis to estimate the color diagnosticity of different foods. We used 20 of the 21 categories (with PRETZEL deleted) and the 11 basic color terms of English (Berlin & Kay, 1969): yellow, white, red, purple, pink,

---

[10] We used a normalized color advantage rather than a simple difference between gray-scale RT and full-color RT because the simple difference may increase with overall RT regardless of prototypicality. Normalization of this sort means that an association between increased color advantage and nonprototypicality might be mediated by overall longer RTs and have nothing to do with increased reliance on color. By normalizing, we are in effect studying the relative rather than the absolute RT advantage conferred by color. Significant correlations exist between prototype score and either the normalized or nonnormalized color advantage.

orange, green, gray, brown, blue, and black.[11] In a questionnaire, subjects were first asked to estimate probabilities of each of the 20 food categories given a specific color. The written instructions included the following:

> On each of the following pages, there appears a color name at the top and a list of 20 food items. Suppose a color photo of one of these foods is concealed. The photo could show a single or group shot of a single food type (e.g., a single coconut or a pile of coconuts), and the food might be whole or in pieces (e.g., a whole coconut or half a coconut showing its interior). You are told only the main color in the picture. Your job is to rate in percentage terms the likelihood of each of the 20 items. The sum of your ratings should add to 100%.
>
> In the absence of any color clue, there is 1 chance in 20 (i.e., a 5% likelihood) for each item. If an item is more likely than other items to have the designated color, you should rate it higher than 5%. If the item is very unlikely to have the designated color compared with other items on the list, rate it less than 5%.

Subjects distributed their numbers unevenly. Given YELLOW, for example, they tended to rate the likelihood of LEMON or BANANA substantially more than 5% and CELERY and PICKLE less than 5%. Each subject completed a set of ratings of this sort for each of the 11 colors. These numbers were taken as the conditional probabilities of food x given color y, $P(F_x \mid C_y)$. The subjects were then asked to assign percentage probabilities to each of the 11 colors (summing to 100%) that represented their likelihood of occurrence across the domain of foods in the study. If the 11 colors were equally likely, subjects were told to assign probabilities of about 1/11 (9%) to each. Again, subjects distributed their numbers unevenly, with higher probabilities for red and green than for blue or black. These numbers were taken as prior probabilities of the colors in food, $P(C_y)$.

The prior probabilities and the conditional probabilities so obtained were used in an information-theory analysis (see the Appendix) to compute the information transmitted by color about food (i.e., entropy reduction concerning the food item, given color). We partitioned the total entropy of food, given color, among the 20 food categories. Categories contributing small amounts of entropy are those for which color is defined to be diagnostic. For each of the 20 food categories, we obtained such a measure of color diagnosticity. Analysis of the RTs from Experiment 2 revealed that color speeded up object recognition more for some food categories than others. We examined the correlation between the RT advantage due to color and our information-theory measure of color diagnosticity for foods.

Results and Discussion

If a subject rated the prior probabilities of the 11 colors as equal, the entropy of the colors, $H(C)$, would be 3.46 bits ($2^{3.46} = 11$). Actual values for the 9 subjects ranged from 2.42 to 3.34 bits, with a mean of 2.91. All subjects indicated that some colors had higher prior probabilities than others.

Subjects were told that the 20 foods were equally likely, corresponding to an entropy, $H(F)$, of 4.32 bits. Once a color is specified, how much uncertainty remains about the type of food? Figure 5A shows mean estimates of $H(F \mid C_j)$ for each of the 11 colors. Given that a food is green, the uncertainty remains high, 3.55 bits. Knowing that a food is green does not restrict the number of possibilities very much—from 20 to about 12 choices (i.e., $2^{3.55} = 11.7$). Purple is the color yielding the least residual uncertainty about the food, just 1.06 bits, which corresponds to slightly more than two equal alternatives.

Computed across all colors, the uncertainty of food, given color, is 2.68 bits, compared with 4.32 bits in the absence of color. The information transmitted by color about food is just the difference, 4.32 − 2.68 = 1.64 bits. The right vertical axis in Figure 5A shows the information transmitted by each of the 11 colors about food, ranging from 0.77 bits for GREEN to 3.26 bits for PURPLE.

As described in the Appendix, we can partition the entropy that remains after color is given, $H(F \mid C)$, into contributions from each of the foods, F_i:

$$H(F \mid C) = \sum_i H(F_i \mid C) \tag{1}$$

The number $H(F_i \mid C)$ is our definition of the color diagnosticity of food, F_i, with low numbers (low entropy) corresponding to high color diagnosticity. Figure 5B shows estimates of color diagnosticity for the 20 foods averaged across 9 subjects. If all 20 foods had equal color diagnosticity, all values in Figure 5B would be 2.68/20 = 0.134 bits. Actual estimates ranged from the least diagnostic values of 0.218 bits for APPLE and 0.182 bits for PEPPER down to the most diagnostic values of 0.076 bits for CARROT and 0.077 bits for LEMON. Surprisingly, APPLE is the food for which color is least diagnostic. This is because APPLE has several symptomatic colors—red, green, yellow, and white (as in sliced apples)—and these are colors shared by many other foods. Color is most diagnostic for CARROT and LEMON. These are foods with a single dominant color. Orange is the dominant color of CARROT and is fairly restrictive in the sense that few of the foods in our experiment were likely to be orange.

These results quantify the informational value of subjects' knowledge about food colors. Subjects may have used such information to improve performance in our reaction time experiments. If so, we would expect foods like CARROT and LEMON that have high color diagnosticity to benefit more from color than foods like APPLE and PEPPER that have low color diagnosticity. The correlation between normalized color advantage (defined in Experiment 4) and color diagnosticity (Figure 5B) was .098, which was not statistically significant.[12]

The results of this experiment suggest that people's explicit knowledge about food color does not account for the advantage of color in real-time object recognition. But four caveats are in order. First, our measures were based on subjective probability estimates. People may not be accu-

---

[11] Twenty categories made the estimation of probabilities easier since even odds equal .05. The selection of the PRETZEL category was arbitrary.

[12] The correlation was based on values from all 84 images. Because color diagnosticity is defined as a category variable, the four images (i.e., different poses) of each food within a category had the same diagnosticity value.

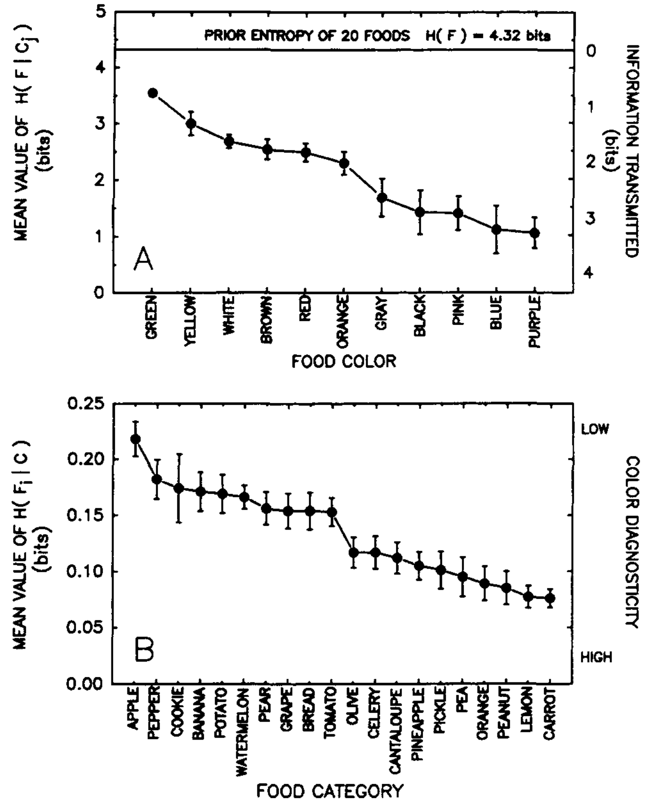

Figure 5. Panel A shows entropies for food given specific colors, $H(F \mid C_j)$, for each of the 11 colors. (The right axis shows the information transmitted by each color. The horizontal dashed line shows the entropy for food in the absence of color information [entropy associated with 20 equally probable choices.] Definitions of the entropies and transmitted information are given in the Appendix.) In Panel B, color diagnosticity is defined to be $H(F_i \mid C)$, which is the portion of $H(F \mid C)$ due to each food F_i. (A high value means low color diagnosticity. The food with lowest color diagnosticity is APPLE, 0.218 bits, and the food with highest color diagnosticity is CARROT, 0.076 bits. The points are mean values for 9 subjects. Error bars show ±1 SE.)

rate in converting their knowledge about food color to numerical probabilities. Second, it is possible that knowledge of food color, manifest in our probability ratings, is distinct from perceptual knowledge of color used in speeded object recognition. For example, the appearance of a particular shade of red on the screen might trigger people's expectations in ways not captured by our analysis. Third, our analysis did not take into account knowledge of color combinations (second-order probabilities). For example, knowing that a target has red and white in it raises the probability of APPLE and reduces the probability of TOMATO. Finally, it is possible that there was some interaction between color diagnosticity and shape diagnosticity. CARROT, for example, was a category with high color diagnosticity. We might expect color to shorten the RT for images in this category. At least one of the images in this category, however, the prototypical single carrot, had a unique shape. For this image at least, strong shape information might have negated the expected color advantage.

Regression analysis. Experiment 4 showed that a perceptual variable, prototypicality, was a predictor, albeit a weak one, of the RT advantage from color. Experiment 5 showed that a cognitive variable, color diagnosticity, was not a predictor.

We used simple and multiple regression to explore the relationship among several variables, both sensory and cognitive, and the RTs measured in Experiment 2. We considered two RT measures associated with each of the 84 images. The overall RT is the mean RT computed across all trials with that image (correct responses only) in the four color-by-blur conditions. The normalized color advantage is equal to the mean gray-scale RT minus the mean full-color RT divided by the average of the gray-scale and full-color RTs.

Cognitive variables. In Experiment 5, we found no significant correlation between color diagnosticity and normalized color advantage, and there was no significant correlation with overall RT.

Word frequency is well known to have an impact on RT (see Paap, McDonald, Schvaneveldt, & Noel, 1987). Our 21 category names ranged from 0.0243 to 76.91 occurrences per million in English text for *cantaloupe* and *bread*, respectively (Carroll, Davies, & Richman, 1971). The standard frequency index (SFI) is a log transformation of word frequency derived by Carroll et al. (1971). We found a significant correlation between overall RT and SFI ($r = -.248$, $p < .05$), but there was no significant correlation with normalized color advantage.

We also considered word length of our category names in letters (3–10) and syllables (1–4) as predictors. Neither of these variables correlated significantly with either overall RT or normalized color advantage.

Sensory–perceptual variables. In Experiment 4, we found significant correlations between prototype score and both overall RT and normalized color advantage. We interpreted this result as support for the view that nonprototypical members of a food category benefit more from color in speeded object recognition.

We found that RT was faster for images with a single food item compared with images with multiple items (e.g., a single carrot vs. a bunch of carrots), ($r = .216$, $p < .05$).[13] This correlation is probably related to the prototype results because best prototypes were often single items. There was no significant correlation with normalized color advantage.

Luminance contrast is certainly suspect as a variable that could interact with color. In the extreme case of an equiluminant image (i.e., no luminance contrast), color variations must be present for an object to be seen. Is the color advantage we observed explained by the low luminance contrast of our images? We considered two definitions of image contrast.

Michelson contrast ranges from 0 to 1 and is defined as $(L_{max} - L_{min})/(L_{max} + L_{min})$, where L_{max} and L_{min} are the maximum and minimum luminances in the image. Our 84 images had high Michelson contrast, ranging from .88 to

---

[13] In this analysis, the predictor was a binary variable; The image had either one or more than one food items.

.999. Michelson contrast was not significantly correlated with either overall RT or normalized color advantage.

The Michelson contrast of an image can be determined by highly localized bright or dark spots (e.g., specularities or shadows.) The root-mean-square (rms) contrast, however, is based on all of the pixels in an image. It is equal to the square root of the luminance variance, normalized by mean luminance, and gives a measure of average luminance modulation around the mean. The rms contrasts of our 84 images ranged from 0.469 to 1.665. The correlation between overall RT and rms contrast was .198 ($p = .07$), but the correlation with normalized color advantage was −.034 and far from significant.

Perhaps signal variation within some limited spatial-frequency band is more closely linked to color advantage than image contrast as a whole. We decomposed our 84 images into Laplacian pyramids (Burt & Adelson, 1983) consisting of six subimages, each with different spatial-frequency content. We computed signal energy for each subimage. There was no significant correlation between signal strength and either overall RT or normalized color advantage for any level of the pyramid (i.e., for any spatial-frequency band).[14]

We considered two measures of image luminance: (a) luminance range within an image, $L_{max} - L_{min}$, and (b) the difference between the mean luminance of an image and the adapting level of 7.94 cd/m^2. Neither of these variables correlated significantly with either overall RT or normalized color advantage.

Finally, we conducted stepwise regression. For normalized color advantage, no variable followed median prototype score in accounting for a significant proportion of the variance. For overall RT, the model included median prototype score, SFI, and word length (in letters). The multiple correlation was .579, compared with $r = .481$ for median prototype score alone.

The color advantage we observed is not accounted for by the cognitive variables we studied or by global measures of image contrast or luminance.

General Discussion

Previous studies (reviewed in the introduction) have been inconclusive with regard to the role of color in object recognition. A possible reason for the discrepancies has been the lack of control over the luminance characteristics of colored and achromatic images. In our study, we compared colored and achromatic images, matched in luminance, and found a decreased RT (and reduced error rate) for the colored images. We conclude that color does improve object recognition.

Our set of objects consisted of familiar foods. Food categories provide a favorable domain for studying the interaction of color and shape because there is natural but limited variation of both attributes within a category. If we had studied a set of man-made objects with highly constrained shape and unlimited color variation (e.g., crayons, cars, key rings, etc.), we might have found less color advantage. There would also have been less opportunity to study the interaction of color and spatial resolution. Alternatively, we could have studied a set of objects in which shape was mostly uninformative and color was necessary for identification (e.g., some species of birds or fish). We probably would have found a larger color advantage, but this choice would have hampered us in studying the interaction of color and spatial resolution. Apart from these considerations, we have no reason to suspect that food objects are perceptually unusual or that our particular selection is unrepresentative.

Our prototype experiment showed that best prototypes exhibit a slightly smaller color advantage than nonprototypical images. The food images that are possibly the most prototypical are those in which shape information is most unique and perceptually prominent. This would be consistent with Biederman and Ju's (1988) view that primary access to object recognition uses structural (geometrical) representations of objects.

One purpose in this study was to test the hypothesis that color improves object recognition more for blurred images. For high-resolution images, texture and shape information may dominate recognition, with color being redundant. When shape and texture are degraded by blur, people may rely more on color cues. Implicit in this hypothesis is the idea that color is a distinctive feature that can be traded off with shape or texture features in object recognition. We did not find the hypothesized interaction between color and spatial resolution and hence have no support for such a feature model.

How could color and shape act additively and noninteractively in object recognition? Perhaps color contributes to an early stage of contour extraction and scene segmentation (De Valois & Switkes, 1983; Walls, 1942). Any role played by color in extracting object contours is likely to rely on low spatial frequencies and hence be relatively insensitive to blur. This is because chromatic contrast sensitivity drops off at high spatial frequencies much faster than luminance contrast sensitivity (Kelly, 1983; Mullen, 1985).

Even if color acts at an early sensory level in object recognition, its role does not appear to be global. Our regression analysis indicated that the effectiveness of color was not linked to global image parameters such as mean luminance, rms contrast, or overall signal strength within spatial-frequency bands. It remains possible, however, that color functions locally in the extraction of features or contours within images.

If color's primary role in object recognition is sensory in nature, it would not be related to people's knowledge of the colors of things. This is consistent with the results of our color-diagnosticity experiment and word-frequency analysis. Although people can use color information in guessing the identity of unseen foods, this knowledge does not account for

---

[14] We also computed a quantity normalized resolution advantage for each image, analogous to normalized color advantage. This quantity represents the RT advantage of unblurred over blurred images. None of the correlations between normalized resolution advantage and band-pass signal strength was significant at the $p = .05$ level: The highest correlation was $r = .212$ ($p = .053$) for the Laplacian band centered at 4 cycles/picture.

the color advantage in rapid visual object recognition. Clearly, one could contrive a task in which color knowledge is necessary for success (e.g., finding a red pickup truck in a parking lot full of blue and green pickups). A task of this sort would be better characterized as search rather than object recognition.

Our interest in color was motivated in part by its possible utility in low vision. In Experiment 3, we found a substantial advantage for colored images over gray-scale images for our low vision subjects (mean difference of 353 ms). We used the low vision data to test a second hypothesis, related to the first, that people with lower acuity would benefit more from color. This hypothesis was not supported. Although people with lower acuity are slower in recognizing objects, the advantage of color is fairly constant across acuities.

Research on low vision reading (cited in the introduction) has shown that there is no practical advantage in coding letters (or backgrounds) with color. But the present results show that color does improve low vision object recognition. One application of this finding is to the prescription of closed-circuit TV (CCTV) magnifiers: These magnifiers have been widely used in recent years to provide high magnification for low vision reading. Until recently, all commercial CCTV magnifiers were monochrome, but several color systems have now appeared. If the primary use of a CCTV magnifier is for reading, there seems to be no advantage to a color display. But if the user wishes to magnify color photos or perform other tasks involving object recognition, our results indicate that a color display may be worthwhile.

References

Berlin, B., & Kay, P. (1969). *Basic color terms: Their universality and evolution.* Berkeley: University of California Press.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review, 94,* 115–147.

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

Burt, P., & Adelson E. (1983). The Laplacian pyramid as a compact image code. *IEEE Transactions on Communication, 31,* 532–540.

Carroll, J. B., Davies, P., & Richman, B. (1971). *Word frequency book.* New York: American Heritage.

Cole, G. R., Stromeyer III, C. F., & Kronauer, R. E. (1990). Visual interactions with luminance and chromatic stimuli. *Journal of the Optical Society of America, A7,* 128–140.

Derrico, J. B., & Buchsbaum, G. (1991). A computational model of spatiochromatic image coding in early vision. *Journal of Visual Communication and Image Representation, 2,* 31–38.

De Valois, K. K., & Switkes, E. (1983). Simultaneous masking interactions between chromatic and luminance gratings. *Journal of the Optical Society of America, 73,* 11–18.

Isenberg, L., Nissen, M. J., & Marchak, L. C. (1990). Attentional processing and the independence of color and orientation. *Journal of Experimental Psychology: Human Perception and Performance, 16,* 869–878.

Jacobs, G. H. (1981). *Comparative color vision.* New York: Academic Press.

Kaiser, P. K. (1972). Colour vision in the legally blind. *Canadian Journal of Ophthalmology, 7,* 302–308.

Kelly, D. H. (1983). Spatiotemporal variation of chromatic and achromatic contrast thresholds. *Journal of the Optical Society of America, 73,* 742–749.

Knoblauch, K., Arditi, A., & Szlyk, J. (1991). Effects of chromatic and luminance contrast on reading. *Journal of the Optical Society of America, A8,* 428–439.

Knowlton, M., & Woo, I. (1989). Assessment of functional color perception. *Journal of Visual Rehabilitation, 3,* 5–22.

Legge, G. E., Parish, D. H., Luebker, A., & Wurm, L. H. (1990). Psychophysics of reading: XI. Comparing color contrast and luminance contrast. *Journal of the Optical Society of America, A7,* 2002–2010.

Legge, G. E., Pelli, D. G., Rubin, G. S., & Schleske, M. M. (1985). Psychophysics of reading: I. Normal vision. *Vision Research, 25,* 239–252.

Legge, G. E., Ross, J. A., Isenberg, L. M., & LaMay, J. A. (1992). Psychophysics of reading: XII. Clinical predictors of low-vision reading speed. *Investigative Ophthalmology and Visual Science, 33,* 677–687.

Legge, G. E., & Rubin, G. S. (1986). Psychophysics of reading: IV. Wavelength effects in normal and low vision. *Journal of the Optical Society of America, A3,* 40–51.

Legge, G. E., Rubin, G. S., Pelli, D. G., & Schleske, M. M. (1985). Psychophysics of reading: II. Low vision. *Vision Research, 25,* 253–266.

Markoff, J. I. (1972). *Target recognition performance with chromatic and achromatic displays* (Research Rep. No. SRM-148). Minneapolis, MN: Honeywell.

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

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.

Paap, K. R., McDonald, J. E., Schvaneveldt, R. W., & Noel, R. W. (1987). Frequency and pronounceability in visually presented naming and lexical decision tasks. In M. Coltheart (Ed)., *Attention and performance XII: The psychology of reading* (pp. 221–244). Hillsdale, NJ: Erlbaum.

Palmer, S., Rosch, E., & Chase P. (1981). Canonical perspective and the perception of objects. In J. Long & A. Baddeley (Eds.), *Attention and performance IX* (pp. 135–151). Hillsdale, NJ: Erlbaum.

Polyak, S. L. (1957). *The vertebrate visual system.* Chicago: University of Chicago Press.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8,* 382–439.

Shannon, C. E., & Weaver, W. (1949). *The mathematical theory of communication.* Urbana: University of Illinois Press.

Sicurella, V. J. (1977). Color contrast as an aid for visually impaired persons. *Journal of Visual Impairment and Blindness, 71,* 252–257.

Steward, J. M., & Cole, B. L. (1989). What do color vision defectives say about everyday tasks? *Optometry and Vision Science, 66,* 288–295.

Tielsch, J. M., Sommer, A., Witt, K., Katz, J., & Royall, R. M. (1990). Blindness and visual impairment in an American urban population: The Baltimore eye survey. *Archives of Ophthalmology, 108,* 286–290.

Walls, G. L. (1942). *The vertebrate eye and its adaptive radiation.* Bloomfield Hills, MI: Cranbrook Institute of Science.

Appendix

Information Theory Analysis of Color Diagnosticity

From Shannon's definition for mutual information T (Shannon & Weaver, 1949), the information transmitted by color about food is $T = H(F) - H(F \mid C)$. Here, $H(F)$ is the entropy of foods in the absence of information about color:

$$H(F) = -\sum_i \mathrm{P}(F_i)\log[\mathrm{P}(F_i)].$$

Subjects were told that the 20 foods were equally likely (consistent with the design of the RT experiment), so all values of $\mathrm{P}(F_i)$ were equal to .05. Therefore, $H(F)$ is the entropy associated with 20 equally likely alternatives, 4.32 bits. (Entropies are computed with $\log_2$ so that values are in bits.)

$H(F \mid C)$ is the conditional entropy of food given color:

$$H(F \mid C) = -\sum_i \sum_j \mathrm{P}(F_i, C_j)\log[\mathrm{P}(F_i \mid C_j)],$$

where the sum is taken across i (20 foods) and j (11 colors), $\mathrm{P}(F_i, C_j)$ is the joint probability of food i and color j, and $\mathrm{P}(F_i \mid C_j)$ is the conditional probability of food i given color j. Because the joint probability $\mathrm{P}(\mathrm{F}_i, C_j) = \mathrm{P}(C_j)\mathrm{P}(F_i \mid \mathrm{C}_j)$, the conditional entropy can be rewritten as follows:

$$H(F \mid C) = -\sum_i \sum_j \mathrm{P}(C_j)\mathrm{P}(F_i \mid C_j)\log[\mathrm{P}(F_i \mid C_j)].$$

The probabilities in this expression are the prior probabilities of color $\mathrm{P}(C_j)$ and the conditional probabilities of food given color $\mathrm{P}(F_i \mid C_j)$, both available from subjects' ratings.

$H(F \mid C)$ is a measure of the uncertainty remaining about the food once the color is known. In our experiment, we estimated a value of 2.68 bits for $H(F \mid C)$ from our data. The difference between this value and the uncertainty about food in the absence of color, 4.32 bits, is the information transmitted about food by color, $T = 4.32 - 2.68 = 1.64$ bits.

We can partition $H(F \mid C)$ among the 11 colors or among the 20 foods as follows:

$$H(F \mid C) = \sum_j \mathrm{P}(C_j)H(F \mid C_j),$$

where

$$H(F \mid C_j) = -\sum_i \mathrm{P}(F_i \mid C_j)\log[\mathrm{P}(F_i \mid C_j]).$$

The 11 entropies $\mathrm{H}(F \mid C_j)$ tell us how much uncertainty is left about the food items when each of the 11 colors C_j is given. Values are shown in Figure 5A. For example, $H(F \mid \mathrm{YELLOW})$ was 3.0 bits (averaged across our 9 subjects).

Finally, we can define color diagnosticity by partitioning $H(F \mid C)$ among the 20 foods:

$$H(F \mid C) = \sum_i H(F_i \mid C),$$

where

$$H(F_i \mid C) = -\sum_j \mathrm{P}(C_j)\mathrm{P}(F_i \mid C_j)\log[\mathrm{P}(F_i \mid C_j)].$$

The 20 numbers $H(F_i \mid C)$ give the separate contributions to the conditional entropy of each of the 20 foods and represent our definition of a food's color diagnosticity.

In our experiment, the conditional entropy $H(F \mid C)$ had an average value of 2.68 bits. If all the food categories contributed equally to this entropy, all values of $H(F_i \mid C)$ would be $2.68/20 = 0.134$ bits. As Figure 5B shows, some foods contributed substantially more entropy than others. The quantity $H(F_i \mid C)$ is a measure of our uncertainty about food F_i, given color. Consider the hypothetical example of a food called BRESH for which color is perfectly diagnostic. Suppose that BRESH is always pink, and no other food is pink. Then, P(BRESH | PINK) is 1.0 and P(BRESH | other color) is 0. Because Plog(1/P) is 0 when P is 0 or 1, all contributions to $H(\mathrm{BRESH} \mid C)$ will be 0. The contribution of BRESH to the conditional entropy is zero.

Consider a more realistic example of a food called DEMP. Suppose that when purple appears, there is a 30% probability of DEMP (Plog[1/P] = .521); when blue appears, there is a 20% probability of DEMP (Plog[1/P] = .464); when any other color appears, there is zero probability of DEMP (Plog[1/P] = 0). Assuming that the 11 colors are equally likely, the value of $H(\mathrm{DEMP} \mid C) = (.521 + .464)/11 = 0.09$ bits. If the prior probabilities of blue and purple had been higher than 1/11, then $H(\mathrm{DEMP} \mid C)$ would have been higher. The color diagnosticity of a food depends on both the prior probabilities of its possible colors and the conditional probability of the food given those colors. In our study, values of $H(F_i \mid C)$ ranged from 0.218 bits for APPLE down to 0.076 bits for CARROT. Higher numbers mean a food contributes more uncertainty, and lower numbers mean that once color is known, the food in question contributes less uncertainty. $H(F_i \mid C)$ is our measure of color diagnosticity.

Received April 9, 1992
Revision received September 3, 1992
Accepted September 4, 1992 ■