Color and size interactions in a real 3D object similarity task

Yazhu Ling
School of Biology (Psychology), University of Newcastle upon Tyne, Newcastle upon Tyne, UK

Anya Hurlbert
School of Biology (Psychology), University of Newcastle upon Tyne, Newcastle upon Tyne, UK

In the natural world, objects are characterized by a variety of attributes, including color and shape. The contributions of these two attributes to object recognition are typically studied independently of each other, yet they are likely to interact in natural tasks. Here we examine whether color and size (a component of shape) interact in a real three-dimensional (3D) object recognition task, using solid domelike objects whose distinct apparent surface colors are independently controlled via spatially restricted illumination from a data projector hidden to the observer. The novel experimental setup preserves natural cues to 3D shape from shading, binocular disparity, motion parallax, and surface texture cues, while also providing the flexibility and ease of computer control. Observers performed three distinct tasks: two unimodal discrimination tasks, and an object similarity task. Depending on the task, the observer was instructed to select the indicated alternative object which was "bigger than," "the same color as," or "most similar to" the designated reference object, all of which varied in both size and color between trials. For both unimodal discrimination tasks, discrimination thresholds for the tested attribute (e.g., color) were increased by differences in the secondary attribute (e.g., size), although this effect was more robust in the color task. For the unimodal size-discrimination task, the strongest effects of the secondary attribute (color) occurred as a perceptual bias, which we call the "saturation-size effect": Objects with more saturated colors appear larger than objects with less saturated colors. In the object similarity task, discrimination thresholds for color or size differences were significantly larger than in the unimodal discrimination tasks. We conclude that color and size interact in determining object similarity, and are effectively analyzed on a coarser scale, due to noise in the similarity estimates of the individual attributes, inter-attribute attentional interactions, or coarser coding of attributes at a "higher" level of object representation.

Keywords: color discrimination, size discrimination, 3D objects, object similarity, color-shape interaction, saturation-size effect, data projector

Introduction

In the natural world, objects are characterized by a variety of attributes, including surface color, texture, and three-dimensional (3D) shape. The contributions of these attributes to object recognition are often studied independently of each other (Biederman & Ju, 1988; Edelman & Poggio, 1991; Humphrey, Goodale, Jakobson, & Servos, 1994; Swain & Ballard, 1990; Yip & Sinha, 2002), yet they are likely to interact in natural tasks. Recent results support this conjecture: For example, the perception of 3D shape influences surface color perception via unconscious mechanisms that take account of mutual illumination (Bloj, Kersten, & Hurlbert, 1999; Doerschner, Boyaci, & Maloney, 2004); binocular disparity cues to 3D shape influence color appearance under illumination changes (Yang & Shevell, 2002); and Stroop-like interferences demonstrate that color and shape interact in object representation (Naor-Raz, Tarr, & Kersten, 2003).

Real objects are unlike the 2D, homogeneously colored, homogeneously bright surfaces that appear in "Mondrian" displays and other simulated images typically used for research in color appearance. The use of such computer-generated and displayed images enables control of color but sacrifices other perceptual cues, such as binocular disparity, 3D luminance shading, specular highlights, mutual illumination, etc. While recently developed radiosity programs enable far more natural and sophisticated simulations of real objects than the typical computer-simulated "Mondrian," the display of 3D simulations on flat computer screens is still inherently problematic even with appropriate binocular disparity cues, because of the conflict between pictorial cues and viewing geometry, screen self-luminosity, and other factors (Hurlbert, 1998). Thus, to study color appearance, color constancy, and recognition of real 3D objects, we have developed a novel setup, which preserves both the advantages conferred by computer-driven control of color as well as the natural binocular and monocular cues to 3D shape (Brainard, Brunt, & Speigle, 1997; Kraft & Brainard, 1999). In this setup, real solid objects are illuminated by a computer-controlled light source, so that the observer is able to adjust the color of either the illumination or object surfaces or both. Compared with the recent attempts using advanced radiosity software to simulate more complex 3D scenes (Delahunt & Brainard, 2004; Doerschner et al., 2004), the setup should allow more real-
istic and varied depictions of surface material, through the variety of objects that can be used.

Although we intend to use this setup for further color research with natural 3D objects (see Movie 1 for an example of how we may change the surface color of a banana while changing the global illumination on nearby fruits and vegetables in a different way), here we demonstrate its use with simple neutral shapes. Our aim in this work is to establish baseline measurements of color discrimination for 3D objects under these viewing conditions and to address three basic questions: How do 3D shape differences influence color discriminability; how do color differences influence shape discriminability; and how do color and shape interact to determine object similarity?

Color speeds naming of an object when it is diagnostic for the shape (e.g., a yellow banana is more quickly identified than a black-and-white line drawing of a banana) or when information about shape is ambiguous or otherwise degraded (Tanaka, Weiskopf, & Williams, 2001; Tanaka & Presnell, 1999). But, as Naor-Raz et al. (2003) emphasize, most past studies of the role of color in object representation have not addressed whether color is an integral attribute of representation. Naor-Raz et al. (2003) conclude from a shape-color Stroop-like effect for pictures of familiar objects that color is intrinsic to object representation. Here we simplify the shape variable to the single component of size, using a collection of solid domes that differ only in radius and height. We also find support for the conclusion that color is intrinsic to object representation, in that color discrimination thresholds increase when the task is to determine similarity of 3D objects and the only relevant variable is size.

Methods

Experimental viewing box

Figure 1 illustrates the design of the experimental viewing box. A front surface mirror (35 cm x 35 cm; Galvoptics float glass) reflects light from a computer-driven DLP data projector (Infocus LP530) into the box. A Pilkington Mirropane (one way observation mirror; 80 cm x 80 cm), fitted diagonally in the center of the box, transmits 38% of the projected light onto the objects arrayed on the bottom of the box, and reflects the remainder to be absorbed by the black velvet at the back. Observers view the image of the illuminated objects through a fixed viewing hole aligned on a perpendicular axis with the center of the observation mirror. The design enables observers to see objects illuminated by the data projector without being made aware of its existence.

The projected image consists of target areas (object “stencils”) segmented from the background, the colors of which are each under independent control. Integral to the setup is the segmentation program, calibrated for the illumination and viewing geometry, which exactly aligns the target areas with the visible surfaces of the individual objects in any given scene, using a pixel-by-pixel mapping between the projected image and a digital image of the scene taken from the observer’s vantage point. By varying the colormap values allocated to a particular stencil, we thereby are able to vary the apparent surface color of a particular object, and likewise for the background. By isolating and altering the colors of only those pixels that are projected onto objects’ surfaces, the segmentation program insures that there is no noticeable “bleeding” of color from objects onto the background.

This setup preserves the natural perceptual cues to shape, such as binocular disparity, motion parallax, luminance gradients, texture gradients, mutual illumination, and specular highlights, while retaining much of the flexibility of traditional CRT rendering systems.
Objects

For the experiments described in this work, we used simple objects made from plaster of Paris. We sculpted a series of hemi-ellipsoidal dome shapes from a hemispherical kitchen mold (radius: 3.5 cm), by slicing off planes of varying thickness from the hemispheres’ bases. Thus, each dome in the series is characterized by its height, from base to peak - as measured with a Vernier Calliper Dial Type (range: 6”/150mm, accuracy: 0.01mm) - and radius, half of the maximal diameter of the circular base. The eight object heights selected for the experiment, labelled from 1 (smallest) to 8 (largest) are listed together with their corresponding radii in Table 1. Figure 2 shows photographs of the kitchen mold, and the largest and smallest objects (objects 8 and 1) used in the experiments.

<table>
<thead>
<tr>
<th>Label</th>
<th>Height (cm)</th>
<th>Radius (cm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.09</td>
<td>3.20</td>
</tr>
<tr>
<td>2</td>
<td>2.19</td>
<td>3.24</td>
</tr>
<tr>
<td>3</td>
<td>2.30</td>
<td>3.29</td>
</tr>
<tr>
<td>4</td>
<td>2.40</td>
<td>3.32</td>
</tr>
<tr>
<td>5</td>
<td>2.51</td>
<td>3.36</td>
</tr>
<tr>
<td>6</td>
<td>2.59</td>
<td>3.38</td>
</tr>
<tr>
<td>7</td>
<td>2.71</td>
<td>3.41</td>
</tr>
<tr>
<td>8</td>
<td>2.80</td>
<td>3.43</td>
</tr>
</tbody>
</table>

Table 1. Experimental object properties. Height is measured from the flat base of the dome to its peak; radius is measured as half of the longest chord of the circular base.

For each object label, several identical objects were produced and painted uniformly with matt white paint.

Projector calibration and color selection

The pattern of projected light was consistent with that formed by an unfocussed light source at a finite distance above the object background; the maximum luminance variation in a specified uniform image was 21.5%, and perceptually undetectable, and the maximum chromaticity variation was insignificant (less than 1%). Over any one object, the luminance variation in the projected image was less than 5%. Between trials within each experimental block, and between blocks, the locations of targeted objects were randomized so that they appeared equally often at each position, thus minimizing any effects of spatial inhomogeneity in luminance.

The Infocus LP530 data projector uses digital light processing (DLP) display technology and, therefore, instead of summing only three primary colors (RGB) to produce color, it adds a further “white” component to increase the overall brightness. Despite attempts (Seime & Hardeberg, 2002; Wyble & Zhang, 2003), no satisfactory analytic method has yet been developed to characterize a four-channel DLP projector, particularly for the inverse model from device independent to device dependent colors. We thus borrowed from methods for four-channel printer characterization and chose the method considered best for that purpose: the tetrahedral interpolation method (Bala, 2002). This method requires dividing the training data set’s color space into several tetrahedrons. To map a particular test color from one space into another, the method locates the tetrahedron in which the color lies, and linearly interpolates between the four corners of the tetrahedron to predict the corresponding location in the other color space.

We confirmed and refined the calibration for each experimental color by measuring the chromaticities of the predicted RGB values (as projected onto object surfaces) with a Photo Research PR-650 spectroradiometer. Where the measured chromaticity did not match the expected value, we altered the RGB values until the best fit was obtained. Table 2 lists all the colors used in the experiments and their peak luminances and chromaticity coordinates. The details of how and why these colors were selected will be described in the following sections.

In the experiments, we employed as our target colors a series of 10 yellowish-greens. We selected these colors first because they lie along the blue-yellow natural daylight axis, and second because the projector performed best in this region of color space. Thus, we were able to obtain sub-threshold color differences between adjacent yellowish-greens. These colors are labeled from YellowGreen 0 to YellowGreen 9 in Table 2, and, as indicated in Table 3, have nearly constant hue and lightness, varying primarily only in saturation.

General experimental procedure

Observers viewed a 4 x 4 array of domes of varying size (at least one individual per size category as labeled in Table 1) and surface color; Figure 3 shows a typical image from the experiments.

On any one trial, only two or three objects (depending on the experiment) were targeted for the task, with the other objects acting as distracters. Small stationary triangu-
lar pointers indicated the targeted objects. The pointer’s
distance from each target object was varied randomly from
trial to trial, within a small range of locations at the bottom
left of each target. The observer used a joystick to move an
additional darker pointer to select one of the targeted ob-
jects, and pressed a joystick button to confirm the choice.
The program then recorded the information for that object
and started the next trial. (Note that the pointers in
Figure 3 have been magnified for clarity in the compressed
figure.) The surface colors of the targeted objects were al-
ways a subset of the YellowGreen colors listed in
Table 2.

The distracters’ colors were randomly assigned so that for
each trial, there would be 4 cyan, 4 pink, and 4 purple ob-
jects in addition to the 4 YellowGreen objects. Thus, the
space-averaged chromaticity of the overall scene on each
trial was near-neutral (CIE Yxy: [11.375, 0.3213, 0.3885]).

On each trial, the locations of the targeted objects were
randomly selected (within the constraints of the experimen-
tal requirements), and the distracter colors were randomly
re-assigned to the remaining objects. After each block of
trials in each experiment (the number of trials depending
on the experiment), the experimenter physically re-arranged
the objects within the array, so that observers were unable
to remember specific size-location associations.

We performed three specific experiments, corresponding
to three distinct tasks: (1) size discrimination; (2) color
discrimination; and (3) object similarity. For the size dis-

Table 2. Luminance and chromaticity coordinates of all the
colors used in all three experimental tasks.

<table>
<thead>
<tr>
<th>Color</th>
<th>CIE Y</th>
<th>CIE x</th>
<th>CIE y</th>
<th>Associated task</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘Background’</td>
<td>10.8</td>
<td>0.32</td>
<td>0.374</td>
<td></td>
</tr>
<tr>
<td>‘Purple’</td>
<td>11.3</td>
<td>0.313</td>
<td>0.343</td>
<td>All</td>
</tr>
<tr>
<td>‘Cyan’</td>
<td>11.3</td>
<td>0.372</td>
<td>0.397</td>
<td>All</td>
</tr>
<tr>
<td>‘Pink’</td>
<td>12.2</td>
<td>0.336</td>
<td>0.357</td>
<td>All</td>
</tr>
<tr>
<td>‘YellowGreen 0’</td>
<td>10.6</td>
<td>0.342</td>
<td>0.432</td>
<td>All</td>
</tr>
<tr>
<td>‘YellowGreen 1’</td>
<td>10.3</td>
<td>0.344</td>
<td>0.435</td>
<td>‘color’, ‘similarity’</td>
</tr>
<tr>
<td>‘YellowGreen 2’</td>
<td>10.4</td>
<td>0.347</td>
<td>0.442</td>
<td></td>
</tr>
<tr>
<td>‘YellowGreen 3’</td>
<td>10.4</td>
<td>0.350</td>
<td>0.448</td>
<td></td>
</tr>
<tr>
<td>‘YellowGreen 4’</td>
<td>10.3</td>
<td>0.352</td>
<td>0.455</td>
<td></td>
</tr>
<tr>
<td>‘YellowGreen 5’</td>
<td>10.3</td>
<td>0.355</td>
<td>0.46</td>
<td></td>
</tr>
<tr>
<td>‘YellowGreen 6’</td>
<td>10.4</td>
<td>0.356</td>
<td>0.467</td>
<td></td>
</tr>
<tr>
<td>‘YellowGreen 7’</td>
<td>11.0</td>
<td>0.361</td>
<td>0.475</td>
<td></td>
</tr>
<tr>
<td>‘YellowGreen 8’</td>
<td>10.7</td>
<td>0.364</td>
<td>0.482</td>
<td>‘size’</td>
</tr>
<tr>
<td>‘YellowGreen 9’</td>
<td>10.1</td>
<td>0.378</td>
<td>0.538</td>
<td></td>
</tr>
</tbody>
</table>

Table 3. Lightness, hue, and saturation values (LHS; in CIE Luv
space) of the 10 target yellowish-green colors.

<table>
<thead>
<tr>
<th>Color</th>
<th>Lightness</th>
<th>Hue</th>
<th>Saturation</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘YellowGreen 0’</td>
<td>99.39</td>
<td>6.12</td>
<td>0.35</td>
</tr>
<tr>
<td>‘YellowGreen 1’</td>
<td>98.18</td>
<td>6.13</td>
<td>0.37</td>
</tr>
<tr>
<td>‘YellowGreen 2’</td>
<td>98.37</td>
<td>6.13</td>
<td>0.41</td>
</tr>
<tr>
<td>‘YellowGreen 3’</td>
<td>98.44</td>
<td>6.14</td>
<td>0.44</td>
</tr>
<tr>
<td>‘YellowGreen 4’</td>
<td>98.11</td>
<td>6.14</td>
<td>0.48</td>
</tr>
<tr>
<td>‘YellowGreen 5’</td>
<td>98.18</td>
<td>6.16</td>
<td>0.51</td>
</tr>
<tr>
<td>‘YellowGreen 6’</td>
<td>98.62</td>
<td>6.13</td>
<td>0.54</td>
</tr>
<tr>
<td>‘YellowGreen 7’</td>
<td>100.43</td>
<td>6.16</td>
<td>0.58</td>
</tr>
<tr>
<td>‘YellowGreen 8’</td>
<td>99.64</td>
<td>6.16</td>
<td>0.61</td>
</tr>
<tr>
<td>‘YellowGreen 9’</td>
<td>97.33</td>
<td>6.08</td>
<td>0.86</td>
</tr>
</tbody>
</table>

Figure 3. A photograph of a typical object array used in the ex-
periments. The pointers indicating the reference object (double
cursor) and its target alternatives (single cursors) have been
darkened and enlarged for illustration.

lar pointers indicated the targeted objects. The pointer’s
distance from each target object was varied randomly from
trial to trial, within a small range of locations at the bottom
left of each target. The observer used a joystick to move an
additional darker pointer to select one of the targeted ob-
jects, and pressed a joystick button to confirm the choice.
The program then recorded the information for that object
and started the next trial. (Note that the pointers in
Figure 3 have been magnified for clarity in the compressed
figure.) The surface colors of the targeted objects were al-
ways a subset of the YellowGreen colors listed in Table 2.
The distracters’ colors were randomly assigned so that for
each trial, there would be 4 cyan, 4 pink, and 4 purple ob-
jects in addition to the 4 YellowGreen objects. Thus, the
space-averaged chromaticity of the overall scene on each
trial was near-neutral (CIE Yxy: [11.375, 0.3213, 0.3885]).

In the size discrimination task, the observer was simply
instructed to choose which of two targeted objects was
“bigger.” On each trial, the targeted objects had the same or
different colors chosen in pairwise combinations from the

Experiment 1: Unimodal size
discrimination task

In the size discrimination task, the observer was simply
instructed to choose which of two targeted objects was
“bigger.” On each trial, the targeted objects had the same or
different colors chosen in pairwise combinations from the
following set of three colors: YellowGreen 0, 8, and 9, with YellowGreen 0 being the least saturated and 9 the most saturated. Four reference sizes were pairwise compared with each of the remaining seven sizes, for each of the five color conditions detailed in Table 4. Note that each reference size and comparison size were represented by least two objects, which between them appeared at all 16 array locations, thereby minimizing any size-location effect.

For each size condition, we tested two reference colors: YellowGreen0 and YellowGreen8. For each reference color and condition, each of the remaining 8 test colors from the full set of 9 listed in Table 3 was tested 16 times. For each reference color and size condition, we fitted a psychometric function to the percentage judged different for each test color.

**Experiment 3: Object similarity task**

In the object similarity task, observers again viewed one reference object with two alternative test objects, but were now instructed to select “the object that is most similar to the reference object.” Observers were not told to attend to any particular feature of the object, but simply to make a choice based on “overall similarity.” Within each session, trials of two different types were randomly mixed. In condition 1, the two alternative test objects had the same color but different sizes, with one alternative having the same size as the reference object. Thus in this condition, the only attribute that could contribute to the choice between the two alternatives was size. In condition 2, the two alternative test objects had the same size but different colors, with one alternative having the same color as the reference object. In this condition, the identical size of the two compared objects provided no extra information, and thus color should provide the only useful cue for the task. In both conditions, the reference object had size 8 and color YellowGreen 8 (i.e., the largest size and most saturated color). All remaining sizes were tested for each of five test colors, and all remaining eight colors were tested for each of four test sizes, with 16 trials per size-color combination.

In both conditions, one of the object attributes should not be relevant for the task, and, therefore, the object similarity task should simplify to the unimodal size or color discrimination task. We therefore plot psychometric functions of the same format as in Experiments 1 and 2, and compare their curves and thresholds to determine whether the assumption of non-interaction between the attributes is valid.

**Observers**

Three males and four females (age range 21-45 years) acted as observers. All had normal color vision as assessed by the Farnsworth-Munsell 100-hue test. Each observer participated in all three experimental tasks.

**Data analysis**

Psychometric functions (logistic) were fitted to the percentage “bigger” (Experiment 1) or “same” (Experiments 2 and 3) responses as a function of object height or color, using the psignifit toolbox, version 2.5.41 for Matlab (see http://bootstrap-software.org/psignifit/), which implements the maximum-likelihood method (Wichmann & Hill, 2001a, 2001b).
2.2 2.4 2.6 2.8
0
0.2
0.4
0.6
0.8
1
0
0.2
0.4
0.6
0.8
1
0
0.2
0.4
0.6
0.8
1

2.2 2.4 2.6 2.8
0
0.2
0.4
0.6
0.8
1
0
0.2
0.4
0.6
0.8
1

2.2 2.4 2.6 2.8
0
0.2
0.4
0.6
0.8
1
0
0.2
0.4
0.6
0.8
1

2.2 2.4 2.6 2.8
0
0.2
0.4
0.6
0.8
1
0
0.2
0.4
0.6
0.8
1

Figure 4. Proportion judged “bigger” for four reference objects in the size discrimination task under five color conditions. “Proportion bigger” data are first averaged over seven observers (solid dots) before logistic functions are fit (solid lines) using psignifit software (see text). Error bars on each dot indicate the standard error of the mean. Each panel’s reference object is as follows: top left – object 1; top right – object 4; bottom left – object 6; bottom right – object 8. Information for each object is listed in Table 1; information for each color condition is listed in Table 4.

Results

Size discrimination task

There were four reference objects and five color conditions in this task. For each color condition, the observer viewed the two indicated objects and selected the “bigger” one. Thus, for each reference object and each color condition, we fit one psychometric function to the proportion judged larger than the reference object, as a function of object height. The psychometric functions fitted to the mean “proportion bigger” data, averaged over all seven observers, are shown in Figure 4.

Each panel of Figure 4 represents a distinct reference object. For all panels, there are systematic shifts in the curves depending on the color condition: Regardless of reference object size, the curves for the conditions in which the reference object is less saturated in color than the test object (cyan and magenta curves) are shifted leftward with respect to those for the conditions in which the reference object is more saturated in color than the test object (red and blue curves). We suggest that these shifts are caused by a perceptual bias to see the more saturated object as larger. This bias is made clear from inspection of the points where the reference and test objects have the same size. For all panels, when the test object height is equal to the reference object height, the observer indicates that the object with the more saturated color is larger on most trials: Thus, the “proportion judged bigger” point differs from the expected 0.5 chance performance. To quantify this effect and demonstrate it more clearly, we define the “perceptual bias” as the difference between the object height corresponding to the 0.5 proportion level on the fitted function (the 0.5 threshold) and the reference object’s height. We display the bias calculated in this way for each reference object size and each color condition in Figure 5.

![Figure 5](https://jov.arvojournals.org/pdfaccess.ashx?url=/data/journals/jov/933508/ on 12/01/2018)
Figure 5 shows that perceptual bias values are always positive for the conditions in which the reference color is more saturated than the test color (red and blue bars). Here when observers are asked to judge whether the test object is larger than reference object, they tend to answer “No.” The effect is reversed for the conditions in which the reference color is less saturated than the test color (magenta and cyan bars), where negative bias values indicate the tendency to answer “Yes.” This group effect occurs within the individual data as well, as summarized by the \( p \) values from one-way ANOVA of the bias values calculated from the psychometric functions fitted to the individual data (not shown). The simplest explanation for these results is that objects with more saturated colors are perceived as larger—we call this the “saturation-size” effect.

We also calculated corrected size discrimination thresholds by subtracting the object height corresponding to the 0.5 “proportion judged bigger” level (the height of subjective equality) from the 0.75 level. This correction removes the effect of the perceptual bias and thus produces a threshold that reflects true discrimination performance on the size task. The results are shown in Figure 6.

Figure 6 shows that for all reference objects except object 1, the bias-corrected discrimination thresholds tend to be smaller when the reference and test objects have the same color (black bar) than when they have different colors, suggesting that size discrimination performance is better when objects have the same color. One-way ANOVA on the individual bias-corrected discrimination thresholds reveals that the effect is statistically significant only for object 8, as indicated by the \( p \) values in Figure 6.

Color discrimination task

In this task, observers viewed a reference object and two test objects simultaneously, one of which always had the same surface color as the reference object. Each observer performed 16 trials per color alternative, for each of two reference colors under each of three size conditions. Figure 7 plots the resulting “proportion-different” data as a function of the test color, measured in units of distance from the neutral origin in CIE xy space (i.e., with saturation

Figure 6. Size discrimination bias-corrected thresholds under five color conditions for four reference objects, computed from the psychometric functions fit to the mean data for seven observers, shown in Figure 4. Note that error bars indicate the bootstrap 95% confidence limits only for the 0.75 thresholds, obtained from psignifit software. \( p \) values indicate the significance level for bias-corrected threshold differences between the color conditions for each object, given by one-way ANOVA [\( F(4,30) \)] on the non-averaged individual fitted values for all seven observers.

Figure 7. Proportion judged different for two reference colors under three size conditions, averaged over seven observers. Error bars on each dot indicate the standard error of the mean. Solid curves represent fitted psychometric functions, obtained by psignifit software. The top panel’s reference color is YellowGreen 0; the bottom panel’s reference color is YellowGreen 8. Information for each color is listed in Table 2 and for each size condition in Table 5.
increasing to the right), averaged over all seven observers. These data were fed into the psignifit engine; the resulting psychometric functions are also shown in Figure 7.

The control curve (black curve; condition 3: reference size object 8 and test size object 1) illustrates the baseline color discrimination performance when no size difference is present. The control curves for both reference colors are steeper than the curves for conditions in which reference and compared objects have different sizes (conditions 1 and 2; red and blue curves). There are no significant differences between the curves for conditions 1 (red curves) and 2 (blue), when the reference/test size pairings are object 8/object 1 and object 1/object 8, respectively. Thus, color discrimination for objects of the same size is significantly better than for objects of different sizes.

Figure 8 summarizes the difference in color discrimination performance between conditions in terms of the average color-difference discrimination thresholds for all seven observers. The discrimination thresholds are calculated as the difference between the reference color and the color corresponding to the 0.75 “proportion-different” level on the fitted function, in units of CIE chromaticity distance from neutral. The smaller the threshold, the better the color discrimination. For both reference colors, discrimination thresholds are significantly smaller for the control condition (black bar).

The control curve (black curve; condition 3: reference size object 8 and test size object 1) illustrates the baseline color discrimination performance when no size difference is present. The control curves for both reference colors are steeper than the curves for conditions in which reference and compared objects have different sizes (conditions 1 and 2; red and blue curves). There are no significant differences between the curves for conditions 1 (red curves) and 2 (blue), when the reference/test size pairings are object 8/object 1 and object 1/object 8, respectively. Thus, color discrimination for objects of the same size is significantly better than for objects of different sizes.

Object similarity task

In the object similarity task, observers viewed one reference object and two alternative test objects on each trial, and were asked to select which alternative was more “similar” to the reference object. The task can be divided into two categories, as described in the methods session: same-color/different-size alternatives, and same-size/different-color alternatives. Although in every testing session, the trial categories were randomly intermixed, here we separate their results in Figure 10 and Figure 12.

Figure 10 shows the results from the same-color category, in which the two test objects had the same color (either YellowGreen 0, 2, 4, 7, or 8) but different sizes, one of which was identical to the reference size (object 8). Each

Figure 9 illustrates the range of individual results on the color discrimination task. Each colored diamond represents the color-difference threshold relative to the control threshold for one observer (i.e., the difference between the discrimination threshold for the condition specified on the x-axis and the threshold for the same-size control condition) (condition 3, reference size object 8 and test size object 8). Negative values thus represent superior discrimination with respect to the control condition, and vice versa. Under both conditions 1 and 2, when the reference object’s size is different from the alternative objects’ sizes, the differences are positive, indicating that for each observer, color discrimination improves when the objects have the same size.

Figure 8. Color discrimination thresholds for three size conditions for each of two reference colors, averaged over seven observers. Error bars indicate the bootstrap 95% confidence limits for the thresholds, obtained from psignifit software.

Figure 9. Color-difference discrimination thresholds for the three size conditions, relative to the same-size control condition (condition 3 in Table 5). Each colored diamond indicates the result for one corresponding observer. The reference color for the top panel is YellowGreen 0, and for the bottom panel, YellowGreen 8.
It is apparent from Figure 10 that the size-discrimination functions do not vary significantly with color (one-way ANOVA confirms that the proportions are not significantly different except at one isolated height each for color pairs 0 and 8, and 7, and 8). In one sense, this result is not surprising: On any given trial of this category, the two alternatives under comparison have the same color, so color should not influence the discrimination task. For all conditions except when the test objects have surface color YellowGreen8, neither alternative is the same color as the reference object, so the similarity should be determined by the size similarity alone. The slight difference visible between Colors 0 and 8 is consistent with the saturation-size effect evident in the unimodal size discrimination task: Here, the larger size of the reference object, which has the more saturated color, should be most discriminable in comparison with the smaller of the two alternative objects, which has the less saturated color. Thus, we would expect the similarity judgment to be sharper than for the control condition, in which the smaller object has the same color as the reference object. But, the sharpening is much less pronounced than in the unimodal size-discrimination task, and not significant.

In fact, because the color-size configurations are exactly the same in the unimodal discrimination and object similarity tasks, we may directly compare the two sets of results. Doing so yields a difference: In the object similarity task, the size-discrimination curves are flattened for each color pair. Figure 11 explicitly shows this difference for the two control conditions in the two tasks. In both conditions, the reference object and alternative objects all have the same color (YellowGreen 8, the most saturated color). The reference object is size 8, the largest, and thus by virtue of color and size, should appear to be the largest in any comparison except with itself. This expectation is borne out in both curves—smaller objects appear more different. (The control curve from Figure 4 has been inverted to represent “proportion-judged-smaller,” and thus accord with the “proportion-judged-different”.)
tion-judged-different” in the similarity task.) But the difference between the discrimination curves for the two tasks shows that it is easier to distinguish the size difference in the unimodal task than in the object similarity task.

Likewise, the slope of the curve for condition 2 for reference object 8 in the unimodal size discrimination task is steeper than for the directly comparable condition (color 0) in the object similarity task. We suggest below that this difference may be due to an interaction between color and size in object representation, which is necessarily accessed in the similarity task but not in the unimodal task. Support for the argument comes from the results for the second category of trial in the object similarity task.

In the second category of trial (same-size/different-color), the reference object was again object 8 with surface color YellowGreen 8. The two alternative test objects had the same size (1, 4, 7, or 8), but differed in color, one alternative having the same color as the reference. Hence, for each size condition, we obtain a color discrimination curve, shown in Figure 12 for the data averaged over all seven observers. Here, the control curve (black) shows the proportion of trials on which the non-matching color was correctly judged as different, for the condition in which the reference and alternative objects all had the same size (object 8).

Similarly to the size-discrimination category discussed above, the color-discrimination curves do not vary significantly with the size of the alternative objects. Yet, again, we may directly compare these results with the results from the analogous conditions in the unimodal color-discrimination task. Figure 13 illustrates the difference between the color discrimination performances for the two control conditions under the two distinct tasks. For each task, the reference and alternative test objects are all of size 8, and the reference color is YellowGreen8. Here, the only difference between the two tasks, on any trial for this particular condition, is in the instructions to the observer: either to choose the object with the same color or to choose the most similar object. Performance is again better for the unimodal task: Its threshold discriminable color difference is less than half that for the object similarity task.

Figure 14 summarizes the difference in average discrimination thresholds, for the two different categories in the object similarity task, relative to their counterparts in the unimodal discrimination tasks. The thresholds in the object similarity task are significantly larger than in the single-attribute discrimination tasks. Table 6 and Table 7, respectively, list the individual observers’ size and color discrimination thresholds for the unimodal discrimination and object similarity tasks. For each attribute, the unimodal threshold is significantly smaller than the corresponding similarity discrimination thresholds, for six out of seven observers. Thus, both for individual observers and on average, discriminations of attribute differences are poorer for the object similarity task.

Before pursuing the implications of this result, we must insure that the difference between the unimodal and bimodal size discrimination thresholds is not due simply to the structural difference between the tasks. Here we used a two-alternative forced-choice (2AFC) task for the unimodal size discrimination task (the observer views only two objects, and chooses the “bigger” one) and an oddity task for the bimodal similarity task (the observer must look at three
The similarity score of each object (size 8, color YG8), and the test object with the highest similarity score is selected. The similarity score of each test object, \( S_{\text{bimodal}} \), is a function of the two distinct similarity scores, \( S_{\text{col}} \) and \( S_{\text{size}} \), which in turn are functions of the unimodal discrimination thresholds, \( \sigma_{\text{col}} \) and \( \sigma_{\text{size}} \), i.e., \( S_{\text{col}} \approx g(\Delta_{\text{col}}, \sigma_{\text{col}}) \) and \( S_{\text{size}} \approx g(\Delta_{\text{size}}, \sigma_{\text{size}}) \), where \( \Delta_{\text{col}} \) represents the difference between the reference and test colors, and \( \Delta_{\text{size}} \) represents the difference between the reference and test sizes. [Note that we do not need to specify \( g \) for the arguments here, but for the purposes of this task, we may define maximal similarity as minimal difference, and therefore model \( g \) as \( 1 - P(\text{"different"}) \), where \( P(\text{"different"}) \) is the probability that the test attribute is judged different. The similarity score here therefore varies from 0 for no similarity to 1 for maximum similarity. \( P \) is in turn related to the proportion judged different \( f^t \) on the unimodal discrimination task, according to the relationship \( P = 2f - 1 \). Under these assumptions, our results exclude a multiplicative model in which \( S_{\text{bimodal}} \) is the product of the two unimodal similarity scores; in that case, the apparent thresholds for color and size discrimination would be reduced relative to their unimodal values. Therefore, we model the total similarity as

\[
S_{\text{bimodal}} = w_{\text{col}} S_{\text{col}} + w_{\text{size}} S_{\text{size}}
\]

where \( w_{\text{col}} (w_{\text{size}}) \) is the weight given to the unimodal color (size) similarity score. This equation assumes that similarity is determined solely on the basis of the two independent attributes, regardless of their combination into a single, coherent object. In general, object similarity may also depend on meta-attributes of the object, and, in this particular case, on an interaction between size and color in the formation of object representations. The combination of size and color may create a set of new descriptors in which the two attributes are inextricably linked, so that—for example—“large-light green” becomes an object quality that confers a distinct identity (say, “apple”) that differs from “small-light green” (say, “grape”) in an object metric that is not simply the conjunction of the color and size scales. While we are not specifically investigating the existence or representation of such meta-attributes, we may formalize the possibility that size and color interact in our object similarity task by postulating the addition of third term in

<table>
<thead>
<tr>
<th>Size discrimination threshold (cm)</th>
<th>UniModal</th>
<th>Similarity</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observer AG</td>
<td>0.012</td>
<td>0.937</td>
<td>0.925*</td>
</tr>
<tr>
<td>Observer ACH</td>
<td>0.367</td>
<td>0.278</td>
<td>-0.089</td>
</tr>
<tr>
<td>Observer JNJ</td>
<td>0.062</td>
<td>0.475</td>
<td>0.413*</td>
</tr>
<tr>
<td>Observer KW</td>
<td>0.012</td>
<td>0.422</td>
<td>0.410*</td>
</tr>
<tr>
<td>Observer LAT</td>
<td>0.012</td>
<td>0.281</td>
<td>0.269*</td>
</tr>
<tr>
<td>Observer SH</td>
<td>0.134</td>
<td>0.325</td>
<td>0.191*</td>
</tr>
<tr>
<td>Observer YL</td>
<td>0.053</td>
<td>0.285</td>
<td>0.232*</td>
</tr>
<tr>
<td>Mean Result</td>
<td>0.074</td>
<td>0.387</td>
<td>0.313*</td>
</tr>
</tbody>
</table>

Table 6. Individual size discrimination thresholds in the unimodal discrimination (bias-corrected) and object similarity tasks, and their difference, for all seven observers. Differences that are significant at the 95% confidence level (as determined by psignifit) are marked with an asterisk.

<table>
<thead>
<tr>
<th>Color discrimination threshold</th>
<th>UniModal</th>
<th>Similarity</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observer AG</td>
<td>0.007</td>
<td>0.017</td>
<td>0.010*</td>
</tr>
<tr>
<td>Observer ACH</td>
<td>0.007</td>
<td>0.021</td>
<td>0.014*</td>
</tr>
<tr>
<td>Observer JNJ</td>
<td>0.013</td>
<td>0.027</td>
<td>0.014*</td>
</tr>
<tr>
<td>Observer KW</td>
<td>0.012</td>
<td>0.026</td>
<td>0.014*</td>
</tr>
<tr>
<td>Observer LAT</td>
<td>0.012</td>
<td>0.028</td>
<td>0.016*</td>
</tr>
<tr>
<td>Observer SH</td>
<td>0.016</td>
<td>0.029</td>
<td>0.013*</td>
</tr>
<tr>
<td>Observer YL</td>
<td>0.011</td>
<td>0.019</td>
<td>0.007</td>
</tr>
<tr>
<td>Mean Result</td>
<td>0.007</td>
<td>0.017</td>
<td>0.010*</td>
</tr>
</tbody>
</table>

Table 7. Color discrimination thresholds in color discrimination and object similarity tasks as well as their differences for all participants. Differences that are significant at the 95% confidence level (as determined by psignifit) are marked with an asterisk.
the above equation: \( w_{\text{obj}} S'_{\text{col} \times \text{size}} = w_{\text{col}} S'_{\text{obj}} \), in which \( S'_{\text{obj}} \) represents an “object” similarity score computed on the hypothetical level of object representation.

Now consider the trials on which color only should determine the similarity. Because the two test objects are always the same size, their size similarity with respect to the reference object, \( S_{\text{size}} \), should be the same. Therefore, if there is no contribution from the higher level \( S'_{\text{obj}} \) term, the sole determining factor in the similarity score, \( S'_{\text{bimodal}} \), should be \( w_{\text{col}} S'_{\text{col}} \), which in turn is determined by \( \sigma_{\text{col}} \).

The difference between the psychometric functions in Figure 13 demonstrates that the unimodal \( \sigma_{\text{col}} \) cannot be the sole determinant of performance on the bimodal task. Thus, it may be that in the bimodal task, \( \sigma_{\text{col}} \) is altered, or that there is a non-zero contribution from either \( S_{\text{size}} \) or \( S'_{\text{obj}} \) or both.

The first explanation, that \( \sigma_{\text{col}} \) alone is altered, is plausible due to the increased attentional demands of the bimodal task. It could be that attention to size in the bimodal task interferes with attention to color, and decreased attention to color increases the color discrimination threshold, \( \sigma_{\text{col}} \). But this explanation implies an interaction between the two attributes in the overlapping of attentional mechanisms. (Note that this implication is not supported by recent evidence for non-interacting attentional streams for distinct attributes within the visual modality [Morrone, Dentl, & Spinelli, 2002].) The second explanation, that there is non-zero contribution from \( S_{\text{size}} \), is also plausible due to inherent noise in the judgment of the attribute values and hence their similarity. For example, if there is noise in the estimate of \( S_{\text{size}} \) it will not necessarily be the same for both test objects, and the similarity score for a dissimilar color may be inappropriately elevated relative to the alternative. In general, inherent noise in the estimate of size similarity will increase the probability that dissimilar colors will be accepted as similar, leading to an apparent increase in color discrimination threshold. But inspection of the difference between the curves in Figure 13 suggests that if noise in size similarity estimation is the main determinant, its effects are not independent of the color difference between the test objects. On every trial under these conditions, the reference and test objects are all of the same size, but for the bimodal task, objects at almost every color difference nonetheless have a higher probability of being judged similar to the reference color than in the unimodal task. This higher probability may result from an increased probability that a test object at that color difference is judged more similar in size to the reference object. If this increase in probability were due to noise in the estimate of size, we would expect it to affect each test object equally, because all are of the same size. But the probability that size similarity is misjudged appears to decrease as the color difference increases (i.e., the difference between the two curves narrows as color difference increases, which is inconsistent with the hypothesis that object similarity is influenced by a constant additive term due to noise in size estimation). This dependence itself suggests an interaction between size and color at some level—the effects of noise in size similarity estimates are moderated by color difference estimates.

Both of these explanations may be augmented by a possible role that the trial history plays within an experimental session. Although on any one trial, size might not need to be considered to make the discrimination, the observer is forced by the demands of the similarity task to consider both it and color on every trial. Thus, size similarity may force acceptance on a color difference that would have been above threshold in the unimodal task; likewise, color similarity may force acceptance on a size difference that would have been unacceptable in the unimodal task. These broadened acceptance thresholds then persist throughout the task, even for trials where sharper differences occur in the other attribute.

Third and last, it may be that there is a significant contribution from the term \( w_{\text{obj}} S'_{\text{obj}} \). It may be that at the level of object representation, where attributes are combined, color representation is coarser than at lower levels. For example, objects of similar size and color may be grouped into categories within which finer differences in either attribute are lost, and the term \( w_{\text{obj}} S'_{\text{obj}} \) becomes critical. From the set of experiments we describe here, we cannot make quantitative distinctions between these explanations. But all of these explanations imply, first, that neither attribute dominates the similarity judgment, and, second, that the two attributes cannot be considered independently in object similarity judgments.

**Discussion**

Our finding that color differences affect the perception of size is perhaps not surprising. The notion that dark colors make one look thinner has long been a tenet of the fashion industry. But the scientific literature itself is thin and conflicting on the topic of color-size interactions. There are reports that perceived saturation increases as the visual angle increases up to 20 deg (Davidoff, 1991), and that larger stimuli (e.g., 30 deg) appear brighter, as well as more saturated in color, than smaller stimuli (e.g., 2 deg) (Jin-Sook, Chang-Shoon, Yon, & Deok-Hyung, 2000), while more recent results suggest that perceived hue and saturation shift non-systematically when stimulus size changes from 10 to 120 deg (Kutas, Bodrog, & Schanda, 2002). Thus while size evidently affects perceived color, the effects are not necessarily predictable. There are also reports of the converse, that color affects perceived size, but only via its dimensions of hue and brightness (Claessen, Overbeeke, & Smets, 1995; Over, 1962). We have shown—possibly for the first time—that perceived saturation affects perceived size. The effect is small but significant, and robust across observers.

The saturation-size effect we observe is most likely not predominantly a 3D size effect. In a separate experiment, two observers (SH and YL) performed the unimodal color
Acknowledgments

YL was supported by a Unilever studentship. We thank Gabi Jordan for the loan of the spectroradiometer.

Commercial relationships: none.
Corresponding author: Yazhu Ling.
Email: yazhu.ling@ncl.ac.uk.
Address: School of Biology (Psychology), University of Newcastle upon Tyne, Newcastle upon Tyne, UK.

References


