Categorical properties of the color term “GOLD”

Gouki Okazawa
Division of Sensory and Cognitive Information, National Institute for Physiological Sciences, Okazaki, Aichi, Japan, & Department of Physiological Sciences, The Graduate University for Advanced Studies (SOKENDAI), Okazaki, Aichi, Japan

Kowa Koida
Division of Sensory and Cognitive Information, National Institute for Physiological Sciences, Okazaki, Aichi, Japan, & Electronics-Inspired Interdisciplinary Research Institute (EIIRIS), Toyohashi University of Technology, Toyohashi, Aichi, Japan

Hidehiko Komatsu
Division of Sensory and Cognitive Information, National Institute for Physiological Sciences, Okazaki, Aichi, Japan, & Department of Physiological Sciences, The Graduate University for Advanced Studies (SOKENDAI), Okazaki, Aichi, Japan

Humans are able to categorize an infinite variety of surface colors into a small number of color terms. Previous studies have shown that 11 basic color terms are commonly used in fully developed languages. These studies usually used flat matte color plates as stimuli, but we can also perceive the colors of glossy surfaces by discounting the effect of the gloss. However, color terms such as GOLD and SILVER are specifically associated with glossy surfaces. In this study, we conducted a categorical color-naming task to examine whether the color terms GOLD and SILVER could be located in a stimulus space defined by combining CIE *xy* chromaticity coordinates and surface reflectance and whether they had categorical properties like ordinary basic color terms. We found that GOLD and SILVER were used for specific ranges of chromaticities with stimuli having large specular reflectances. Moreover, the strengths of the categorical properties, as assessed using measures of consistency, consensus, and reaction time, were comparable to those of the basic color terms, indicating that GOLD and SILVER are categorical color terms specifically associated with glossy surfaces. This also indicates that humans do not always discount surface gloss to identify colors but can utilize this information to categorize colors.

Keywords: color vision, visual cognition, categorization, material perception

Citation: Okazawa, G., Koida, K., & Komatsu, H. (2011). Categorical properties of the color term “GOLD”. *Journal of Vision, 11*(8):4, 1–19, http://www.journalofvision.org/content/11/8/4, doi:10.1167/11.8.4.

Introduction

In our daily life, we encounter objects with an infinite variety of surface colors. Although we can discriminate millions of different colors, it would be impossible to identify or memorize all these colors. Instead, we categorize similar colors into a small number of groups and describe them using a limited number of color terms. Each language has its own color terms, but common properties exist across different languages. Through their anthropological surveys, Berlin and Kay (1969) found that there are 11 color terms (basic color terms) that are used commonly in fully developed languages. Psychophysical color-naming experiments have also shown that the basic color terms are more frequently used than other color terms (Boynton & Olson, 1987; Guest & Van Laar, 2000). For example, Boynton and Olson (1987) instructed subjects to name the colors of OSA color plates and, using several response measures, showed that the basic color terms are consistently and commonly used across subjects. This categorical nature of color terms is known to affect various aspects of color perception, including color discrimination (Özgen & Davies, 2002), color memory (Uchikawa & Shinoda, 1996), and visual search (Yokoi & Uchikawa, 2005).

Color categories are affected by the mode of their appearance. Two modes have been identified in color

doi: 10.1167/11.8.4 Received February 24, 2011; published July 7, 2011 ISSN 1534-7362 © ARVO

perception: aperture color mode and surface color mode. Humans categorize colors differently in the two modes, even when their colorimetric values are identical (Uchikawa, Uchikawa, & Boynton, 1989). Whereas the aperture color mode is not directly related to object colors (unrelated color), the surface color mode is for colors of object surfaces. For example, brown and black are named only in the surface color mode, not in the aperture mode (Uchikawa et al., 1989). In addition, many studies have shown that color categories are rather constant under illuminant changes (Hansen, Walter, & Gegenfurtner, 2007; Olkkonen, Hansen, & Gegenfurtner, 2009; Olkkonen, Witzel, Hansen, & Gegenfurtner, 2010; Troost & de Weert, 1991). These results suggest that the process of categorical color naming may involve estimating surface reflectance.

The surface reflectance of most natural and man-made objects can be characterized using two components: diffuse reflection and specular reflection. For materials having both components, such as plastic or ceramic, diffuse reflection carries the material color and specular reflection carries the illumination color. To characterize the color of these materials, it would seem necessary to discount the specular reflection in order to isolate the color contained in the diffuse reflection. Indeed, several recent studies examining the effect of gloss on surface color have shown that humans perceive surface colors by discounting the specular reflection (Giesel & Gegenfurtner, 2010; Todd, Norman, & Mingolla, 2004; Xiao & Brainard, 2008). However, the situation is quite different for materials such as metal, which have only specular reflection. For these materials, specular reflection carries the color unique to the material. Therefore, to identify the material, it would seem necessary to utilize the signals carried by the specular reflection rather than discounting it. There are color terms closely associated with metallic materials, such as GOLD and SILVER. (Note that when we refer to the color term, the word will be spelled in capital letters, like "GOLD." When we refer to the material, the word will be spelled in lower case letters, like "gold.") These color terms are in marked contrast to the ordinary basic color terms, which are used irrespective of the material and glossiness.

The purpose of this study was to determine whether GOLD and SILVER can be localized in a stimulus space defined by combining color and surface reflectance and to test whether these color terms have categorical properties comparable to those of the previously established basic color terms. Localization of GOLD and SILVER as a stable category in the color/reflectance space would provide strong evidence that color categorization can be based on the estimation of surface reflectance properties, such as gloss. Furthermore, it would also indicate that humans categorize surface colors not only by discounting the glossiness but also by using glossiness as a unique property of surfaces.

To address these questions, we generated a set of visual stimuli that was defined in a three-dimensional space similar to the one previously used to examine the perception of GOLD (Nishizawa, 2007; Nishizawa, Segawa, & Uchikawa, 2006), which was composed of a combination of the CIE *xy* chromaticity coordinates and diffuse/specular reflectance. Using these stimuli, we determined the ranges of GOLD, SILVER, and COPPER within this stimulus space using the color-naming procedure previously used to identify the categories of the basic color terms. We also tested whether GOLD, SILVER, and COPPER have categorical properties like ordinary basic color terms. To estimate the categorical properties, we employed several previously used response measures, including consistency within individuals, consensus among subjects, and reaction time. The results show that GOLD and SILVER, but not COPPER, are clearly localized in the stimulus space we generated and that their categorical properties are comparable to those of ordinary basic color terms. This indicates that humans use surface reflectance properties like gloss to categorize surface colors.

Experiment 1

In this experiment, we tested whether the color terms GOLD and SILVER possess categorical properties such that they are stably used for specific ranges of chromaticity and surface reflectance. To address this question, we generated a set of object images whose color and glossiness were systematically manipulated. These images were presented to the subjects, who then named the color of each stimulus (categorical color-naming procedure).

Methods

Apparatus and stimuli

The experiment was conducted in a dark room, and the subjects were allowed to adapt for about 3 min before the experiment was started. Each subject's head was fixed using a chin rest, and the stimuli were presented on a display (Totoku CCL254i2; 1–750 cd/m^2) placed in front of the subject at a viewing distance of 86 cm. The stimuli were presented on a black background (1 cd/m^2, x = 0.296, y = 0.307), and they subtended a visual angle of about 3.5 deg. We calibrated the luminance and chromaticity of the display using a colorimeter (CS200, KONICA MINOLTA).

The stimuli consisted of images of three-dimensional (3D) objects generated using RADIANCE rendering software (Ward, 1994). Scene configurations in RADIANCE have three components: a 3D shape model, illumination, and a surface reflectance model. The 3D shape was created (Figure 1A) by the experimenters using LightWave Modeler (NewTek). The illumination was Eucalyptus Globe obtained from Paul Debevec's high dynamic range illumination database (Debevec, 1998). For the surface reflectance model, we used the Ward–Duer model, which is thought to be physically realistic (Ward, 1994). In the Ward–Duer model, images are rendered by setting three parameters: specular reflectance, diffuse reflectance, and roughness. For the specular and diffuse reflectances, we used the following four combinations of values:

$$(\text{specular, diffuse}) = (1.0, 0.0), (0.8, 0.2), (0.4, 0.6), (0.0, 1.0). \quad (1)$$

Higher specular reflectance and lower diffuse reflectance increased the glossiness of the image. We varied diffuse reflectance and specular reflectance simultaneously because appearance of gloss is affected by both specular and diffuse reflectances (Ferwerda, Pellacini, & Greenberg, 2001) and metallic appearance of an object depends on its specular/diffuse ratio rather than the specular reflectance itself (Motoyoshi, Nishizawa, & Uchikawa, 2007). Hereafter, for simplicity the value of the specular reflectance will be used to indicate each combination (e.g., a specular reflectance of 1.0 will be used for (specular, diffuse) = (1.0, 0.0)). Roughness defines the spread of specular highlights, and a higher roughness value yields a blurred reflection. In the present experiment, the roughness value was fixed at zero, so the rendered objects had sharp highlights, unless the specular reflectance was zero.

In the color-naming experiment, we examined the relationship between the chromaticity of a given image and the color terms used to describe it. However, not only did the images generated using the aforementioned procedures have varying luminances across each image, they also had varying chromaticities because the natural scene used for the rendering contained varying chromaticities. Because each stimulus in the color-naming experiment should have a clearly defined color, we gave all the pixels in each stimulus the same chromaticity, and the luminance values were unchanged. We selected 67 chromaticity coordinates centered at yellow and including orange, green, and white (Figure 1B). The luminance of each pixel was mapped between 1 and 300 cd/m^2 so that the images were within the dynamic range of the display. The mean luminances of the reflectances (specular = 1.0, 0.8, 0.4, 0.0) were 63.2, 72.4, 86.6, and 93.9 cd/m^2, respectively. Their luminance histograms are shown in Figure 1C.

Subjects

Six Japanese subjects participated in the experiment: three males and three females, aged 23 to 36 years. Two of the subjects were authors in this study. All had normal or corrected-to-normal visual acuity and had trichromatic color vision, as determined using Ishihara color plates. We obtained informed consent from all subjects. This research was approved by the Ethics Committee for Human Research of National Institute for Physiological Sciences.

Procedure

The subjects were instructed to select the color of the presented stimulus from 15 color terms including the

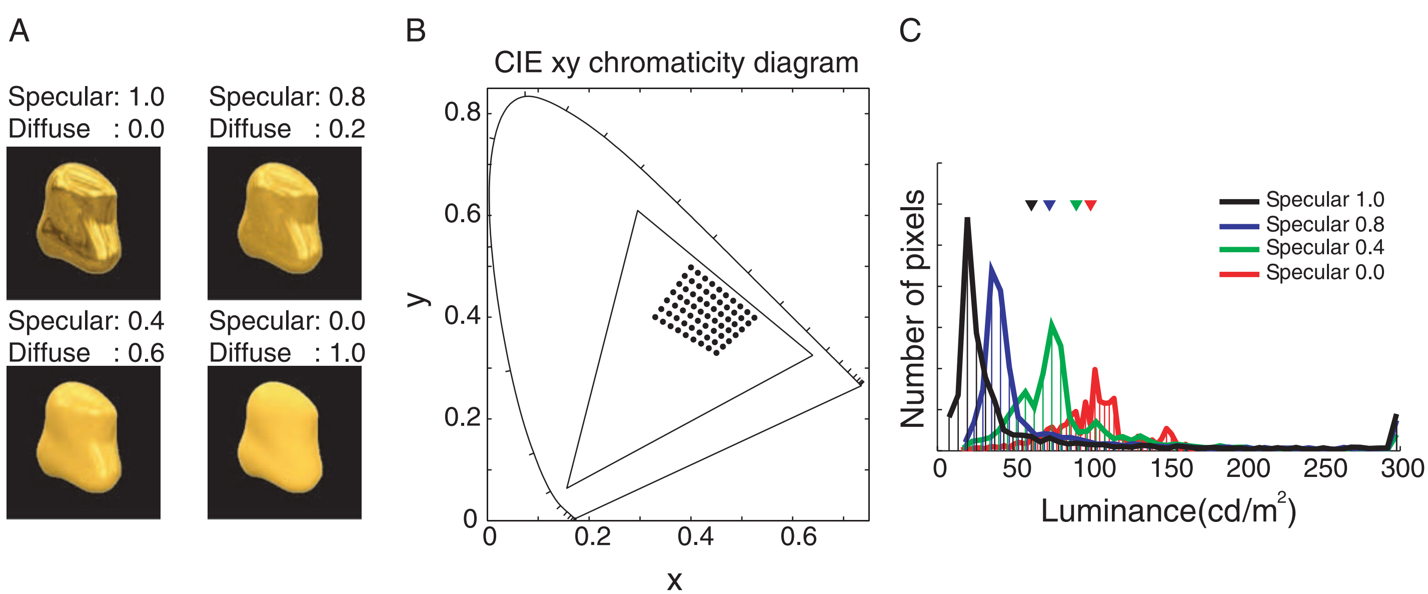

Figure 1. Stimulus configuration in Experiment 1. (A) Examples of the stimuli used in the experiment. Four combinations of surface reflectance parameters (diffuse and specular reflectances) were used to render the images. The values above each image indicate the reflectance parameters. Note that the color of the printed image should be different from the actual stimulus image presented on the display. (B) The 67 CIE *xy* chromaticity coordinates used in the experiment. Each dot represents the chromaticity of a stimulus. (C) Luminance histogram of the image at each reflectance level plotted in different colors. Inverted triangles indicate the mean luminance.

11 basic color terms (green, yellow, orange, pink, brown, red, purple, blue, white, gray, and black), beige, GOLD, SILVER, and COPPER. The Japanese terms for these 15 colors are *midori, ki, daidai, momo, cha, aka, murasaki, shiro, hai, kuro, hada, kin, gin,* and *dou*. The subjects usually answered the name in Japanese, but sometimes they used the English terms (e.g., pink, orange) because they are widely used in Japan. In the analysis, these answers were treated the same as the corresponding Japanese terms. We included *hada* (beige) because Japanese frequently use this color term (Uchikawa & Boynton, 1987), especially in the color range used in this experiment. Moreover, beige is not a basic color term; thus, beige could become a benchmark for evaluating the categorical nature of GOLD.

The subjects were allowed to use one or two color terms. If they selected two terms, they were asked to rank them. We regarded the first as the named color term and used it for the analyses. In each trial, one stimulus was presented at the center of the display, after which the subject said the color term into a microphone and then pressed a button. The stimulus was turned off when the button was pressed. Reaction time was measured as the time interval between stimulus onset and button press. There were no limits to the reaction time. The intertrial interval between the button press and the next stimulus presentation was 1 s. Each stimulus was presented twice to each subject, and there were a total of 536 trials (67 chromaticities $\times$ 4 reflectances $\times$ 2). Before making measurements for analysis, the subjects performed about 50 trials as training, which confirmed that the subjects could correctly perform the task.

Data analysis

To evaluate the categorical properties of the named color terms, we calculated two indexes, consistency and consensus, which have been used previously in studies of categorical color naming (Boynton & Olson, 1990; Uchikawa & Boynton, 1987). Consistency is a measure of how stably a color term was used for the same stimulus by each individual subject and was defined as follows for each color term:

$$\text{Consistency} = (\text{the number of stimuli to which each subject repeated the same color term in both trials} \times 2/\text{total number of trials in which that color term was used}) \times 100(\%). \quad (2)$$

The responses of all the subjects were combined before the calculation so that a single consistency value would be obtained. For example, if "red" was used in 100 trials across all the subjects, and among them 50 trials were for the same stimulus by the same subject (i.e., 2 times for 25 stimuli), the consistency of red was 50%. Consensus is a measure of the degree to which a color term was used for the same stimulus across subjects. For a given color term, this value was computed for each stimulus to which that color term was applied and defined as follows:

$$\text{Consensus} = (\text{the number of trials in which a given color term was used for one stimulus}/\text{the number of all trials for the stimulus}) \times 100(\%). \quad (3)$$

The denominator (the number of all trials for one stimulus) should be 12 because, in this experiment, the six subjects responded two times to each stimulus. Because consistency is a within-subjects measure, it can sometimes occur that each subject consistently names a stimulus, but the named color terms differ among the subjects. Consensus, on the other hand, is a measure of agreement among the subjects. Both consistency and consensus were computed based on the data obtained using all four reflectance levels.

In this experiment, the subjects were required to name the color of the 67 colored stimuli in the set (Figure 1B) at four different reflectances. To quantify the similarity of the named color terms used for the stimuli between different reflectance pairs, we defined a Category Correlation Coefficient (CR). For each stimulus with a particular reflectance, we counted the number of trials in which the subjects used each color term and generated a matrix (67 colors $\times$ 15 color terms) whose element was the number of trials and their value ranges from 0 to 12. We created this matrix for each reflectance and then calculated the Pearson's Correlation Coefficient between the matrix of one reflectance and that of another reflectance. The calculated values were defined as the CRs, which ranged from 0 to 1, with values close to 1 indicating that the naming results were similar across different reflectances.

For statistical tests, we calculated the probability that the patterns of color naming were the same between two different reflectances as a null hypothesis. To assess the statistical significance of the difference, we computed the probability distribution of the CR for the color naming of the stimuli with each of the two reflectances. To do this, we randomly divided the naming data for each reflectance into two halves, computed the CR (auto-CR) between each half, and repeated the same procedure 20,000 times. We then examined the statistical significance of the actual CR between the two reflectances by testing whether it was within 5% of the lower end of the auto-CR distribution. We selected the distribution whose average was smaller, which made the test of significance more conservative.

Results

Color naming for stimuli with a specular reflectance of 0.0

Before considering the naming of stimuli with high specular reflectance, we analyzed the naming of matte stimuli with a specular reflectance of 0.0. Previous studies of categorical color naming employed matte color plates, and we anticipated results similar to those obtained previously with these stimuli. Figure 2A shows the average of the naming across all six subjects for stimuli with a specular reflectance of 0.0. Each stimulus was named 12 times (6 subjects responded 2 times each). Each symbol represents a color term that was used in more than 7 of the 12 responses (more than 50% of trials). The size of each symbol represents the percent of trials in which a given color term was used.

For stimuli with a specular reflectance of 0.0, basic color terms (yellow, green, orange, pink, and white) and beige were frequently used. The use of these terms and their chromaticity ranges were similar to earlier studies employing Japanese subjects (Uchikawa & Boynton, 1987). To quantitatively compare the naming of images with a specular reflectance of 0.0 with the naming of uniform color plates, we conducted another color-naming experiment with the same subjects using circular uniform color patches (Figure 2B). The luminance of the patches was matched to the mean luminance of the 3D shape stimuli with a specular reflectance = 0.0 (93.9 cd/m^2). The results were very similar to the naming of the matte stimuli, and the CR (see Methods section) between the two conditions was 0.937 ($p > 0.5$).

Color naming using GOLD, SILVER, and COPPER

Figure 3 shows the results of color naming for stimuli with a specular reflectance of 0.4, 0.8, or 1.0. The conventions are the same as in Figure 2. Table A1 shows the number of times each color term was used and the mean reaction time (RT) for each stimulus. Only the results for saturated colors are shown because the number of stimuli was large.

For the stimuli with a specular reflectance of 0.4, the results were nearly the same as those obtained with a specular reflectance of 0.0, and the distribution was not significantly different (CR = 0.949, $p > 0.5$). However, when the specular reflectance was increased further, quite different results emerged. When stimuli with a specular reflectance of 0.8 were used, GOLD was named in more than 50% of trials for stimuli between the yellow and orange regions (CR for data from images with a specular reflectance of 0.0 = 0.782, $p < 0.05$), and for the stimuli with a specular reflectance of 1.0, GOLD was named in a even larger region (CR for data from images with a specular reflectance of 0.8 = 0.721, $p < 0.001$). Some stimuli within this region were called GOLD in a large proportion of trials (over 10/12 trials), which was highest (11/12) for stimuli at $x = 0.469$, $y = 0.443$. Conversely, there was no region that was named yellow or orange when stimuli had a specular reflectance of 1.0 (CR from data with a specular reflectance of 0.0 = 0.357, $p < 0.001$).

When the specular reflectance was 1.0, there were also regions where the stimuli were frequently named SILVER or COPPER (Figure 3). The region for SILVER was centered at chromaticity coordinates around $x = 0.38$, $y = 0.39$,

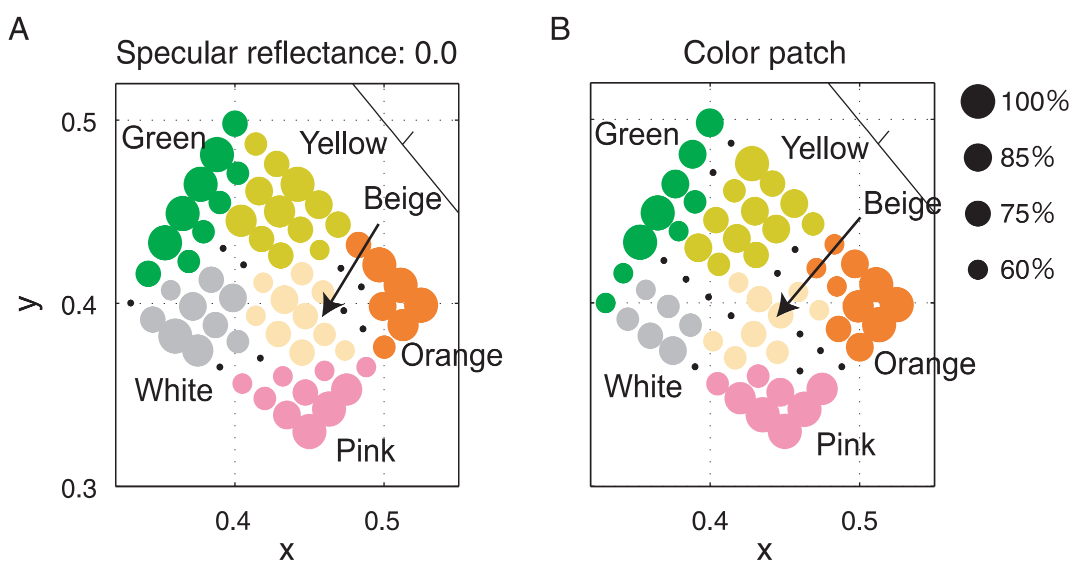

Figure 2. Results from the categorical color-naming task for stimuli with (A) a specular reflectance of 0.0 and (B) uniform color patches plotted on CIE *xy* chromaticity diagrams. Each symbol represents a color term that was named in more than 50% of trials for the stimulus with the corresponding chromaticity. Symbol sizes represent the percentage of trials in which the color term was named. Black dots represent stimuli for which no color term was used in more than 50% of trials.

which roughly corresponded to where stimuli with lower specular reflectances were named white. The region for COPPER ranged between $x = 0.45$ and 0.52 and $y = 0.38$, where stimuli with lower reflectances were named beige or orange.

These results indicate that color terms specific for metallic materials, namely, GOLD, SILVER, and COPPER, emerge as the specular reflectance of the stimuli increase, while use of the basic color terms, such as yellow, orange, and white, correspondingly decline. Stimuli with a specular reflectance of 0.8 were at the transition; consequently, the color naming was unstable and did not reach a high consensus level (>50%) for many chromaticity coordinates. On the other hand, green and pink were named stably across all specular reflectance conditions. Other basic color terms, including gray, blue, brown, black, and purple, were seldom used for the chromaticity range covered in the present study.

Chromaticity of GOLD

In the results summarized above, GOLD was used to name highly specular stimuli within a specific range of chromaticity. This prompted us to ask how the chromaticity of GOLD is related to other basic color terms, especially yellow and orange, whose chromaticity ranges clearly overlap that of GOLD for stimuli with lower reflectances. To address that question, we examined the relationships among these three color terms, comparing the dominant wavelengths for yellow, orange, and GOLD. The dominant wavelength is the point of monochromatic light at which an extrapolated line connecting the stimulus point and the white point ($x = y = 1/3$) intersect with the spectrum locus. Figure 4 shows the average of the dominant wavelengths of all the stimuli that were named yellow, orange, or GOLD by each subject. We used all the trials independently, such that if one stimulus was named twice with the same color term, the chromaticity of the stimulus was counted twice. A horizontal broken line indicates the wavelength of unique yellow (576.8 nm). This value was calculated by averaging the wavelengths of unique yellow determined in several previous studies, which were reviewed in Table 1 of Ayama, Nakatsue, and Kaiser (1987).

The averaged wavelength for yellow (575.9 nm) was very close to the unique yellow (576.8 nm), although there was variation across the subjects. The dominant wavelength of orange ranged from 586 to 590 nm (mean: 588.6 nm) and its variation across the subjects correlated with that of yellow ($r = 0.93$, $p < 0.01$).

The dominant wavelength for GOLD ranged from 577 to 584 nm (mean: 580.7 nm), which was located between yellow and orange for all subjects and significantly differed from both (yellow, $p < 0.01$; orange, $p < 0.001$, t-test uncorrected for multiple comparisons), and its variation correlated with those of both yellow ($r = 0.92$, $p < 0.01$) and orange ($r = 0.86$, $p < 0.05$). This means that although the chromaticity of GOLD is similar to those of yellow and orange, it is distinct from these color terms.

The chromaticity of SILVER was nearly the same as that of white. The mean chromaticity of the stimuli named SILVER across all subjects was $x = 0.384$ and $y = 0.394$, while that of white was $x = 0.384$ and $y = 0.397$.

Consistency and consensus for the naming of GOLD

The results described so far indicate that GOLD was used to name highly specular object images within a specific range of chromaticity. Our next question was whether the color term GOLD has categorical properties like those observed with the basic color terms. To examine this, we calculated two indexes of categorical

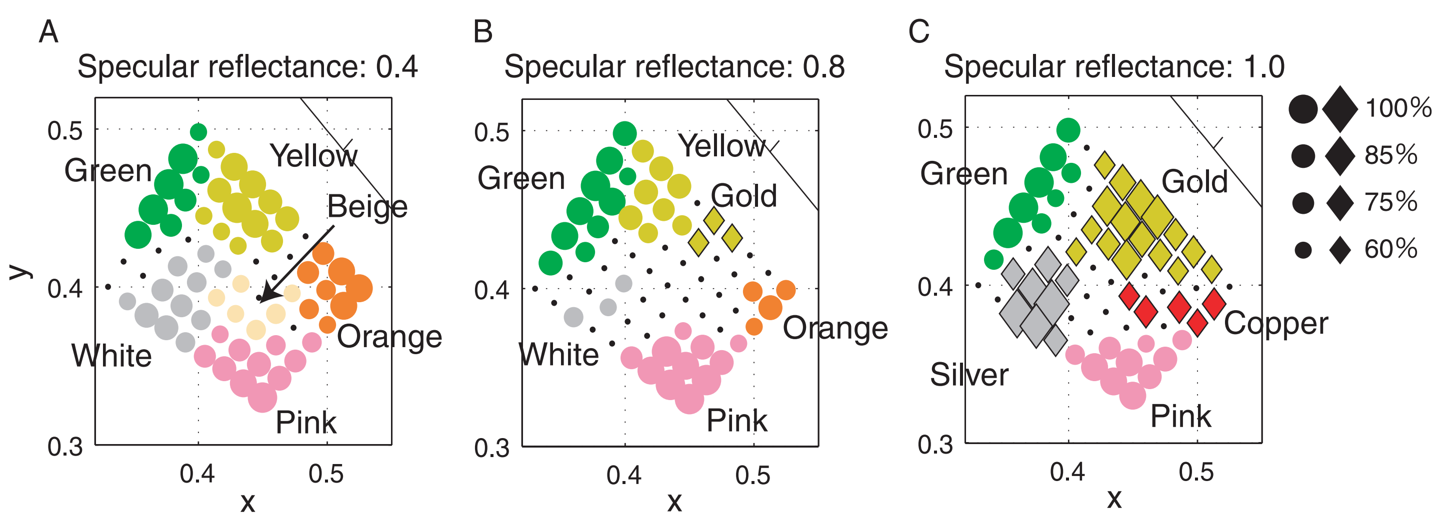

Figure 3. Results from categorical color-naming tasks in which stimuli had specular reflectances of (A) 0.4, (B) 0.8, and (C) 1.0. The conventions are the same as in Figure 2.

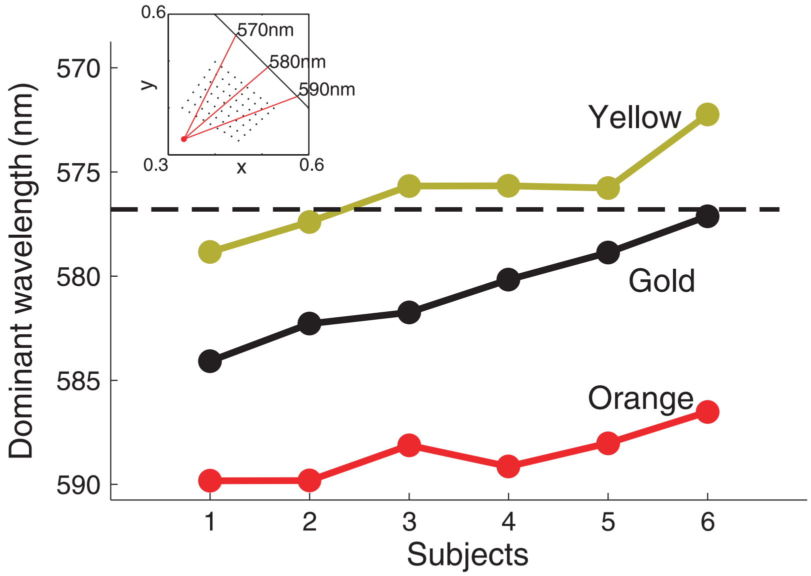

Figure 4. Mean dominant wavelengths of all the stimuli named yellow, GOLD, and orange by each subject. If one stimulus was named twice with the same color term, the chromaticity of the stimulus was counted twice. The dotted horizontal black line indicates the dominant wavelength of unique yellow calculated by averaging the unique yellows determined in several earlier studies (reviewed in Table 1 of Ayama et al., 1987). The inset shows lines connecting the white point ($x = y = 1/3$) with the points of monochromatic light (570, 580, and 590 nm) together with the chromaticity coordinates of the stimuli.

properties, consistency and consensus (see Methods section), which have been used previously in studies of the categorical nature of the basic color terms (Uchikawa & Boynton, 1987). With these indexes, we compared the categorical properties of GOLD, SILVER, and COPPER to those of the basic color terms used in the naming experiment described above.

We first considered the consistency of the naming of each color term. Figure 5A shows the consistency of five basic color terms (yellow, green, orange, pink, and white) that were frequently applied to the stimuli used in this experiment. The values of the consistency ranged from 60% to 82%. Although the range of chromaticity examined in this experiment did not contain the category

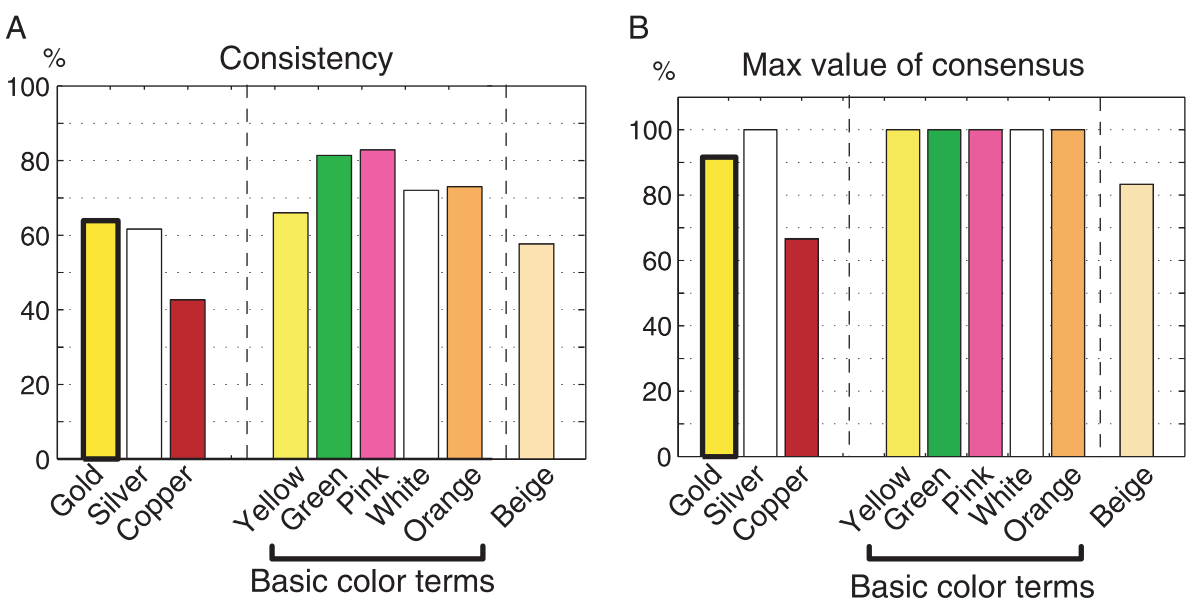

Figure 5. (A) Consistency and (B) the maximum value of consensus for each color term named in the categorical color-naming task.

center of green or white, the consistencies of these color terms are very similar to those reported in previous studies (61% to 81% in Uchikawa & Boynton, 1987). The category center of yellow was within the chromaticity range we used, and its consistency was 66%, which is also similar to that reported previously (61% in Uchikawa & Boynton, 1987). The consistency of beige (58%), a non-basic color term, was similar to those of the basic color terms, though it was slightly lower than that previously observed (67% in Uchikawa & Boynton, 1987). Most importantly, the consistencies of GOLD (64%) and SILVER (62%) were comparable to, albeit slightly lower than, those of the basic color terms. By contrast, the consistency of COPPER (43%) was clearly lower than those of GOLD, SILVER, and the basic color terms.

We next considered consensus, which is another index of categorical properties (Figure 5B). In previous studies, the maximum consensus rate among all stimuli for each color term was adopted as a measure of the categorical nature of that color term (Boynton, MacLaury, & Uchikawa, 1989; Uchikawa & Boynton, 1987). We conducted the same analysis.

The maximum consensus rates of the basic color terms were 100%, which is consistent with earlier reports showing that consensus tends to be very high for the basic color terms (85% to 100% in Uchikawa & Boynton, 1987). The maximum consensus rate for beige was slightly lower than for the basic color terms, which is consistent with the results obtained previously with Japanese subjects (80% in Uchikawa & Boynton, 1987). The maximum consensus rate for GOLD was 91%, which was close to those for the basic color terms (though slightly lower) but was higher than that for beige (83%). The maximum consensus rate for SILVER was also high (100%), whereas the consensus for COPPER (67%) was lower than those for GOLD, SILVER, and the basic color terms. These results are consistent with those for consistency and indicate that GOLD and SILVER have categorical properties comparable in strength to those for basic color terms.

Reaction time

Reaction time is another psychophysical measure previously used to assess the categorical nature of the basic color terms (Boynton & Olson, 1990). Those studies showed that the basic color terms were named with faster reaction times than other color terms. To test whether the reaction time for the naming of GOLD is as fast as for the basic color terms, we calculated the mean reaction time for the stimulus with the maximum consensus rate for each color term. Figure 6A shows the average of all six observers' trials. The average reaction time for the basic color terms (green, yellow, pink, orange, and white) was 2.3 ± 1.5 s. The average reaction time for GOLD was 2.0 ± 1.2 s, which was not significantly different from that for the basic color terms ($p > 0.05$, *t*-test uncorrected for multiple comparisons). The average reaction times for SILVER (2.0 ± 0.8 s) and beige (2.0 ± 1.2 s) also did not significantly differ from that for the basic color terms ($p > 0.05$, *t*-test uncorrected

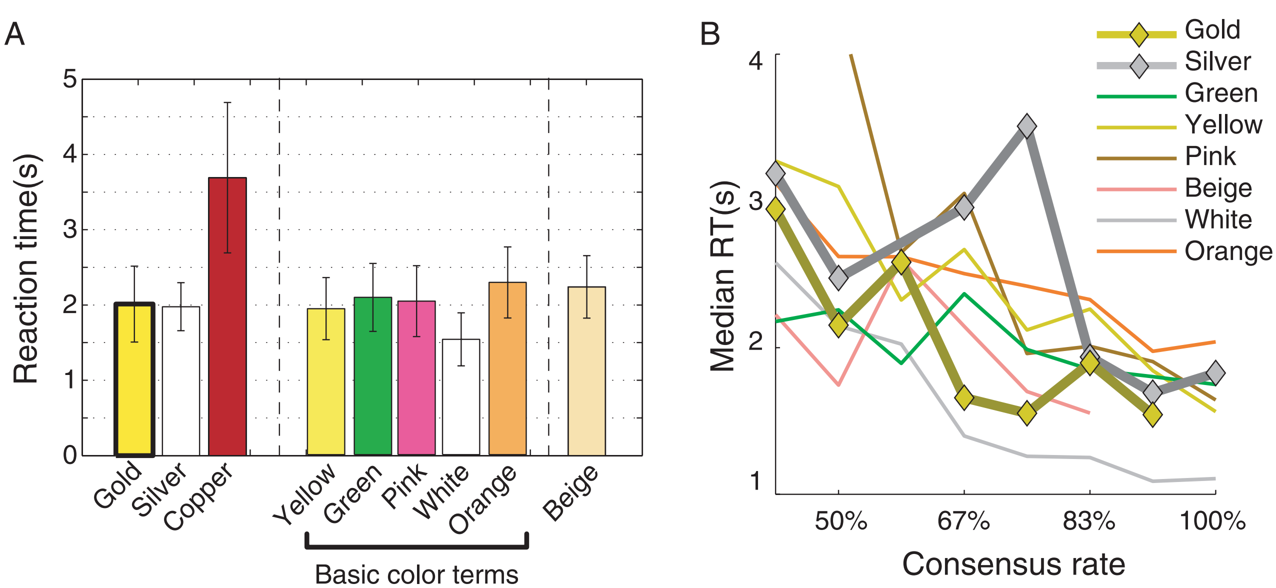

Figure 6. (A) Mean reaction times across all subjects and trials for the stimulus that had the maximum consensus rate for each color term. Lower values represent faster responses. The error bars represent the standard error of the mean. (B) Median reaction times across all subjects and trials in which the indicated color term was used for the stimuli with each consensus rate. When there was more than one stimulus with the corresponding consensus rate, the reaction times of these stimuli were averaged. Only the terms whose maximum consensus rate was larger than 70% are shown.

for multiple comparisons). By contrast, the average reaction time for COPPER (3.7 ± 2.4 s) was significantly longer than that for the basic color terms ($p < 0.0001$, t-test uncorrected for multiple comparisons).

The shorter reaction times indicate that subjects did not hesitate to select the color terms. We suggest that reaction times are shorter for stimuli located near the category center than for those located near the border between categories. To test whether this is the case, we plotted the median reaction time against the consensus rates for each stimulus (Figure 6B) because a stimulus at a category center should have a higher consensus rate than one near the border between categories. When we averaged the reaction times from trials in which a given color term was used with the same consensus rate, the reaction times and consensus rates for the basic color terms (50% to 100%) were negatively correlated ($r = -0.736$ to -0.945, $p < 0.05$). The same was true for GOLD ($r = -0.815$, $p < 0.05$), which indicates that stimuli near the category center of the basic color terms and GOLD were named faster. Because the data for SILVER was noisy due to the small number of trials at the 75% consensus rate, the correlation was not significant ($r = -0.658$, $p = 0.10$).

Color confusion

The subjects sometimes used two different color terms for the same stimulus in different trials. For example, when a stimulus was greenish yellow, subjects used green in one trial and yellow in another trial. This could reflect similarity or affinity between two color terms. In an earlier study, the number of stimuli named with a pair of color terms (color confusion) was counted as a measure of the relationship between the two terms (Boynton & Olson, 1987), and such analysis of color confusion has revealed several important properties of color term relationships (e.g., opponent color terms were not confused). To examine the relationships between GOLD, SILVER, COPPER, and other color terms, we calculated the color confusion rate for each subject as follows:

$$\text{color confusion rate} = \text{number of stimuli named using two color terms in different trials by the same subject} / \text{number of stimuli named using one of the two color terms in at least one trial} * 100(\%). \tag{4}$$

For example, subject 1 used yellow and green in at least one trial for 36 and 48 stimuli, respectively, and used both color terms for 5 stimuli in different trials. Therefore, the confusion rate of subject 1 between yellow and green is $5 \,/\, (36 + 48 - 5) * 100 \div 6\%$.

Figure 7A shows the confusion rates for each pair of color terms averaged across subjects. The number in each box represents the confusion rate (%) between the two terms, and the gray level of each box represents the

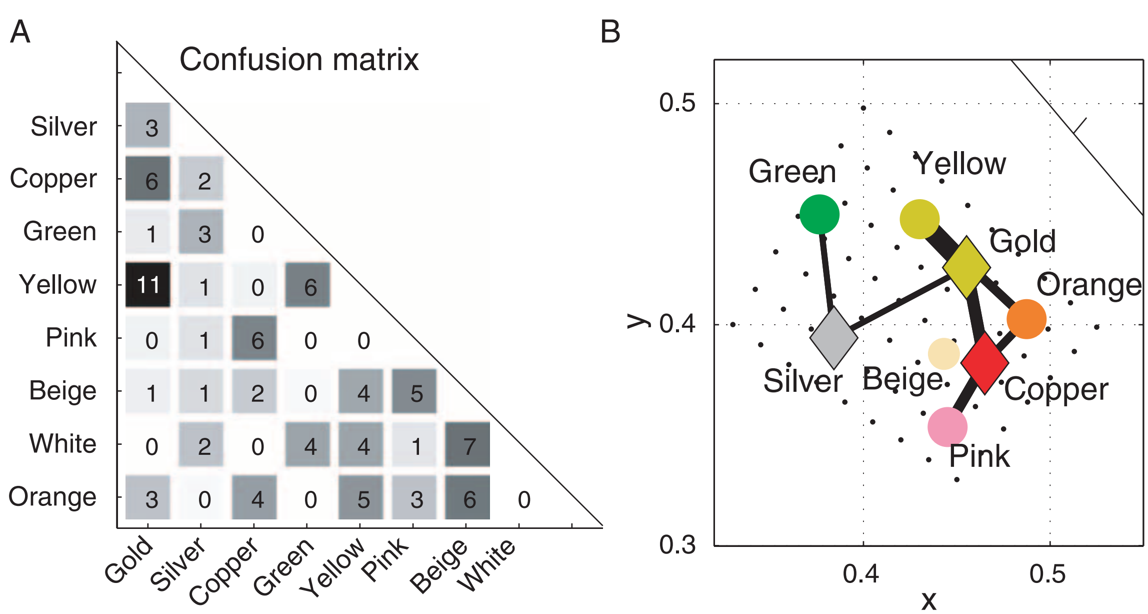

Figure 7. (A) The confusion rates for pairs of color terms. Each boxed number represents a confusion rate (percent of trials in which two color terms were applied to the same stimulus). The gray level of each box represents the magnitude of the confusion rate; pairs of color terms having higher confusion rates are colored darker. (B) Pairs of color terms that had a confusion rate of >3% were connected by a line. The width of the line represents the confusion rate. The symbols are the same in Figures 2 and 3. The position of each symbol represents the average CIE *xy* chromaticity coordinates for all the stimuli named by the corresponding color term.

magnitude of the confusion rate. In Figure 7B, lines connect pairs of color terms that were confused with metallic color terms (GOLD, SILVER, COPPER) more than 3% of the time. GOLD had high confusion rates (>3%) with SILVER, COPPER, yellow, and orange. By contrast, yellow had high confusion rates with GOLD, green, white, orange, and beige, and this pattern was distinct from that of GOLD. Notably, although the chromaticity range of GOLD was located adjacent to that of green when the specular reflectance was 1.0 (Figure 3), its confusion rate with green was small (1%). This suggests that there is a tight boundary between the GOLD and green categories. There was a high rate of confusion between SILVER and GOLD or green, and SILVER exhibited some degree of confusion with many other color terms. There was also a high rate of confusion between COPPER and GOLD, orange, or pink. These patterns of confusion were not observed with any other color terms, which suggests that there are specific relationships between basic color and metallic color terms (e.g., repulsion between green and GOLD). Otherwise, the results were generally as expected in the chromatic ranges of GOLD, SILVER, and COPPER and were consistent with the idea that the chromatic position of GOLD is distinct from any of the basic color terms.

Effect of pixel-level shuffling of stimuli

The results summarized so far indicate that stimuli with the same chromaticity can be named differently, depending upon the surface reflectances that affect various image features, such as the luminance distribution and spatial pattern. These differences in color naming could be attributable to differences in the image features, which raises the question: what image features are specifically related to the naming of metallic color terms, such as GOLD and SILVER? To test whether the luminance distribution is sufficient for the naming of metallic color terms, we randomly rearranged the pixels of images with a specular reflectance of 1.0 (pixel-level shuffling; Figure 8A). The pixel-level shuffling changed the spatial pattern, while maintaining the luminance histogram. We then conducted the color-naming task with the same six subjects using the same procedures described above.

Figure 8B (left panel) shows the results using the same conventions used in Figure 2. The pixel-level shuffling substantially reduced the number of times the name GOLD was applied to a stimulus, and there was no chromaticity named GOLD in more than 50% of trials (CR for the stimuli with a specular reflectance of 1.0 was 0.657). In addition, the naming was unstable, and for 33 of 67 chromaticities (the small black dots in Figure 8B, left panel), no color terms were named in more than 50% of trials.

To more closely scrutinize how these shuffled images were named, Figure 8B (right panel) shows the color terms used in more than 30% of trials. With some chromaticities, more than one color term was used in more than 30% of trials, and the symbols for different terms overlap. Note that in this figure the symbols for GOLD do not form a specific region. Among all four reflectance levels tested in this study, the overall pattern of naming was most similar to the naming of images with a specular reflectance of 0.8 (CR = 0.896), and a moderate correlation was seen with low reflectance images (CR =

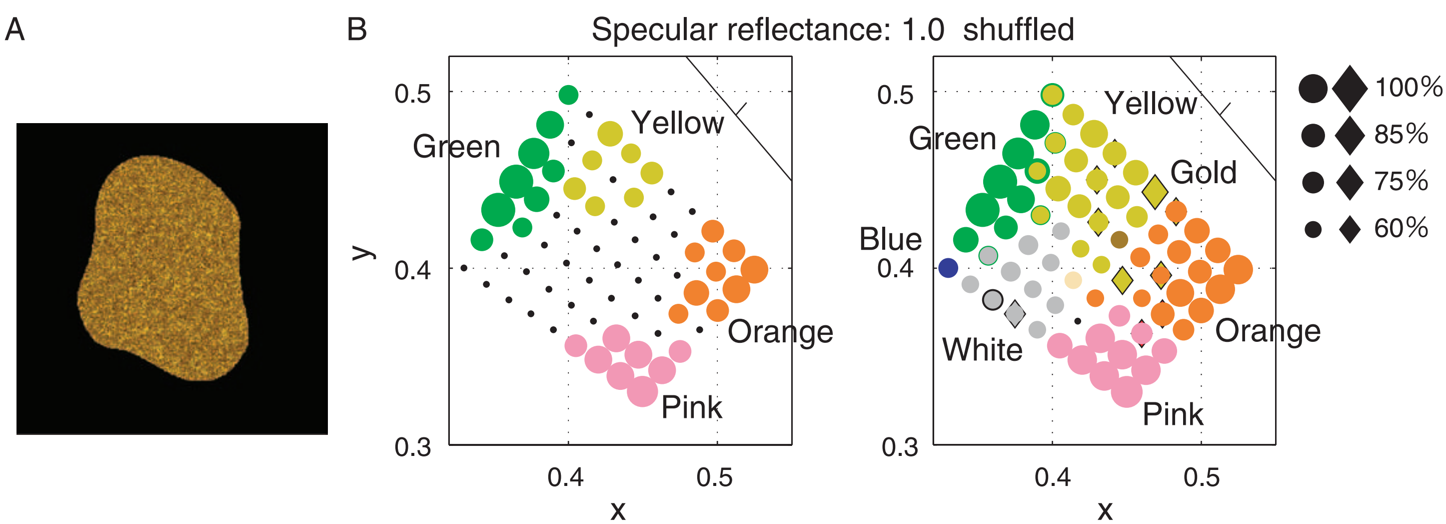

Figure 8. Results of categorical color-naming tasks after pixel-level shuffling of stimuli with a specular reflectance of 1.0. (A) Example of a pixel-level shuffled stimulus. (B) Results of the color-naming tasks. In the left panel, the conventions are the same as in Figure 3. In the right panel, color terms that were used in more than 30% of trials for each stimulus are plotted. The symbols overlap each other when more than one color term was used in over 30% of trials for a stimulus.

0.81 and 0.77 for stimuli with a specular reflectance of 0.4 or 0.0, respectively). This pattern was most dissimilar from that for the original image (reflectance 1.0, CR = 0.66). These findings indicate that the luminance distribution acts together with the specific chromaticity to some degree as a cue, but it is not sufficient for the perception of GOLD, which suggests that the spatial pattern of an image is an important element.

Experiment 2

In Experiment 1, we showed that GOLD had categorical properties and was stably named for CG images with specific ranges of chromaticity and high specular reflectance. These CG images were created using a fixed 3D shape under a fixed illumination condition. It is thus important to know whether the properties associated with GOLD naming are invariant across different shape/illumination conditions. In Experiment 2, we created images with various 3D shapes and illuminations and conducted a GOLD rating task to examine whether these factors affect the perception of GOLD.

Methods

We conducted a GOLD rating task for stimuli with various shapes and illuminations. The stimuli had constant chromaticities, as in Experiment 1. We used four 3D shapes (the shape used in Experiment 1 and three other shapes created using CG) and five illuminations obtained from the Paul Debevec database (Eucalyptus, Campus, Beach, Building, and St. Peter's). All combinations of these four shapes and five illuminations (total: 20 combinations) were rendered with a fixed reflectance parameter (specular reflectance = 1.0, diffuse reflectance = 0.0, roughness = 0). Each rendered image was colored with one of the seven chromaticities shown as blue symbols in Figure 9 ($20 \times 7 = 140$ stimuli).

In addition to the CG images, we prepared five photographs of metallic objects to test whether the perception changes with more natural 3D structures and illuminations. One object was a pendant photographed in our laboratory; the other four were obtained from web sites. These five images of real objects were colored with the same seven chromaticities (total: 35 images) used for the CG images. Note that, for the images of real objects, there was only one type of illumination (i.e., photographed conditions). In total, 175 stimuli were created, and each was presented on a display against a black background (1 cd/m^2, $x = 0.296$, $y = 0.307$).

The subjects' task was to judge whether the color of each stimulus looked like GOLD by rating them from 5 (looks like GOLD) to 1 (does not look like GOLD). Each stimulus

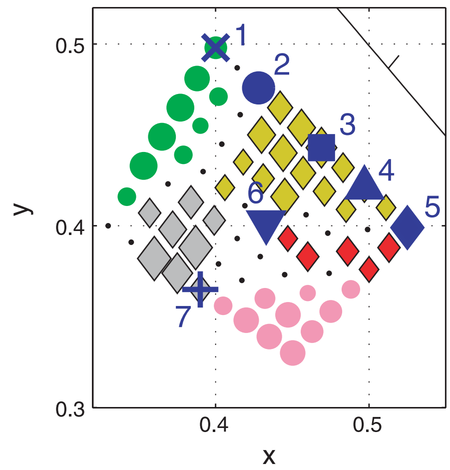

Figure 9. The seven chromaticity coordinates used in Experiment 2 are shown as blue symbols. The symbols are overlaid on the naming data for stimuli with a specular reflectance of 1.0 in Experiment 1 (same as Figure 3C).

was presented three times to each subject. Other experimental conditions were the same as in Experiment 1, and the same six subjects participated.

Results

Figure 10A shows the average GOLD rating for each shape. All the ratings for each shape were averaged across the different illuminations and across subjects. Figure 10B shows the average GOLD rating for each illumination. In this figure, all of the ratings under each illumination were averaged across the different shapes and subjects. The rightmost points in Figure 10B indicate the average GOLD ratings for the five photographed images.

For the CG images, we conducted a 3-way ANOVA with shape, illumination, and color as main factors and assessed the effect of each factor and their interaction. We found that the GOLD rating was strongly influenced by color ($F(6,700) = 147.58$, $p < 0.000001$), as expected, and that shape and illumination affected the rating only weakly (shape: $F(3,700) = 2.82$, $p = 0.0383$; illumination: $(4,700) = 3.92$, $p = 0.0037$). Notably, there were no significant interaction between the color and shape factors or between the color and illumination factors ($p > 0.5$), indicating that the color selectivity for GOLD perception was unaffected by object shape or illumination. Moreover, the colors rated with higher values (color #3) were located near the category center for GOLD observed in Experiment 1 (Figure 3C), indicating that the ratings are in good agreement with the results of Experiment 1.

For the photographed images (the rightmost points in Figure 10B), we conducted a 2-way ANOVA with color and object as main factors and found that the GOLD rating strongly depended on color ($F(6,175) = 29.82$, $p <$ 0.0001) but only weakly on object ($F(4,175) = 2.47$, $p =$

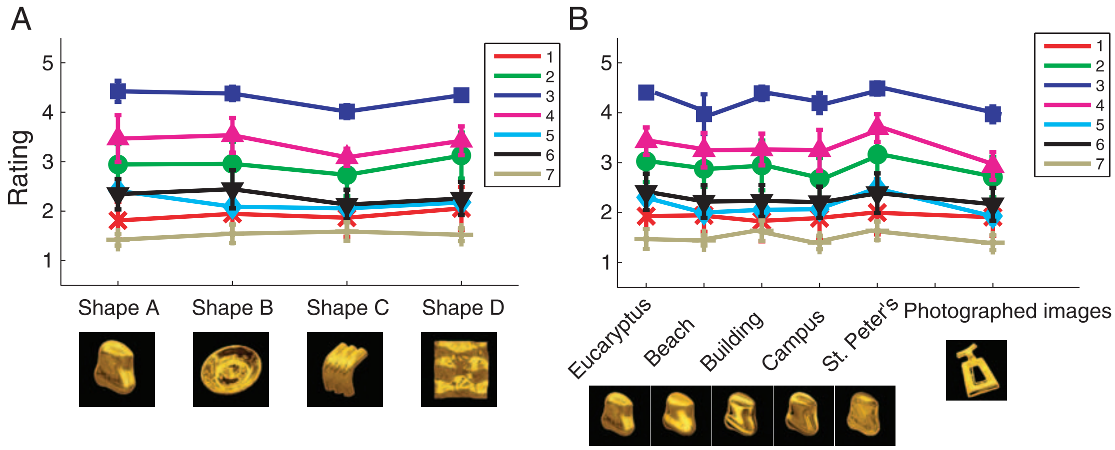

Figure 10. Results of GOLD rating for stimuli (A) with different shapes or (B) under different illuminations. The vertical axis represents the rated value ranging from 5 (looks like GOLD) to 1 (does not look like GOLD). Each colored line corresponds to a chromaticity whose symbol is the same as in Figure 9. Error bars indicate the standard error of the mean. (A) Average GOLD ratings plotted against the indicated shapes. The ratings for each shape were averaged across all illuminations and subjects. (B) Average GOLD ratings plotted against the different illuminations. The ratings to each illumination were averaged across all shapes and subjects. The rightmost data are the average GOLD ratings for five photographed images of real gold objects.

0.0467), and that there was no interaction ($p > 0.5$). When we compared the average ratings of the CG and photographed images, we found no significant difference for any color ($p > 0.05$, *t*-test uncorrected for multiple comparisons). Thus, the results obtained with the photographed images were essentially the same as obtained with the CG images, indicating that colors perceived as GOLD are not affected by the reality of the images.

Experiment 3

In Experiment 1, the subjects did not freely name colors of stimuli but were required to select from 15 color terms given by the experimenters. Therefore, there remain possibilities that color terms that are specifically used for glossy stimuli besides GOLD and SILVER may exist. Furthermore, one might argue that the measures of categorical properties, namely, consistency and consensus, could be biased by restricting the available color terms. To clarify these points, in Experiment 3, we conducted a free color-naming experiment in which subjects can freely give color names. In addition, as colors of stimuli used in Experiment 1 was restricted within the narrow ranges in the color space, we used stimuli with wider ranges of colors in Experiment 3 to increase the chance to recruit any other color terms.

Methods

We used stimuli with specular reflectances of 0.0 and 1.0 and having the same 3D shape with that of Experiment 1. The luminance patterns of the stimuli were the same as those of Experiment 1. As the stimulus color, we selected 49 chromaticity coordinates that covered the display gamut as wide as possible with the required luminance range (illustrated in Figure 11). The total number of stimuli was 98. The procedures for stimulus presentation and the display setting were the same as those in Experiment 1.

Five subjects participated; we recruited three new subjects for Experiment 3 who were naive to the purpose of the experiment and the other two subjects had participated in Experiment 1. They were instructed to name the color of presented stimuli freely while keeping several rules listed below:

1. The term should be monolexic and should be named without adjectives.
2. The term should represent a color of a stimulus as a whole.
3. The term can also be the name of an object as long as it is not the name of a particular product or item.

The first rule was employed in a previous study of free color naming (Boynton & Olson, 1987), whereas the second and third rules were not listed in the previous study. We included the second rule because our stimuli were not uniform and the third rule because some commonly used color terms, such as orange, sky, water, etc., are also the name of real objects.

Each stimulus was presented twice for each subject. The total number of trials was 196. We calculated the consistency and maximum consensus of each color term using the same equation described in the Methods section of Experiment 1.

Color term (*Japanese term*)	No. of trials used	No. of subjects used	Consistency (%)	Consensus (%)
Green	187 (105)	5 (3)	84 (80)	100 (100)
Pink	178 (123)	5 (3)	87 (88)	100 (100)
Yellow	94 (47)	5 (3)	57 (55)	90 (83)
SILVER	84 (60)	5 (3)	86 (87)	100 (100)
Water (*mizu*)	83 (47)	5 (3)	82 (77)	90 (100)
Orange	73 (56)	5 (3)	77 (79)	100 (100)
GOLD	60 (33)	5 (3)	63 (67)	90 (100)
White	59 (33)	5 (3)	81 (79)	90 (83)
Beige	34 (12)	4 (2)	65 (50)	70 (67)
Gray	27 (27)	3 (3)	67 (67)	40 (67)
Purple	20 (3)	3 (1)	80 (67)	60 (33)
COPPER	19 (0)	2 (0)	63 (0)	30 (0)
Blue	15 (5)	3 (2)	27 (0)	30 (17)
Brown	15 (15)	2 (2)	40 (40)	30 (50)
Brownish green (*uguisu*)	12 (12)	2 (2)	33 (33)	30 (50)
Lemon	7 (7)	2 (2)	29 (29)	30 (50)
Chrome	6 (0)	1 (0)	100 (0)	20 (0)
Cream	5 (5)	1 (1)	40 (40)	20 (33)
Egg (*tamago*)	5 (5)	1 (1)	40 (40)	20 (33)
Turquoise	2 (0)	1 (0)	0 (0)	10 (0)
Bright yellow (*yamabuki*)	2 (2)	1 (1)	0 (0)	10 (17)
Red	1 (1)	1 (1)	0 (0)	10 (17)
Sky (*sora*)	1 (1)	1 (1)	0 (0)	10 (17)
Emerald	1 (1)	1 (1)	0 (0)	10 (17)
Aqua	1 (0)	1 (0)	0 (0)	10 (0)
Brass (*sinchu*)	1 (0)	1 (0)	0 (0)	10 (0)

Table 1. Color terms and their number of trials used, number of subjects used, consistency, and consensus observed in Experiment 3 (free color-naming task). *Note*: Values with all five subjects (values with three naive subjects).

Results

In total, twenty-six color terms were used by five subjects. All the terms are listed in Table 1. Of these, ten ordinary basic color terms, GOLD, SILVER, COPPER, and beige were the color terms used in Experiment 1. Other color terms included *mizu, uguisu, lemon, chrome, cream, tamago, turquoise, yamabuki, sora, emerald, aqua,* and *sinchu* in Japanese. The corresponding English terms are water, brownish green, lemon, chrome, cream, egg, turquoise, bright yellow, sky, emerald, aqua, and brass, respectively. One subject used a term *kogane* that has the equivalent meaning and interchangeable with *kin* (Japanese term used in Experiment 1 that corresponds to GOLD in English) and we treated *kogane* the same as *kin* in the following analysis. The results using the data of three naive subjects are provided in parentheses in each column in Table 1. As the results of the new subjects and experienced subjects were very similar (Table 1), we combined them in the following analysis.

Figures 11A and 11B show the color terms used for more than 50% of trials for each stimulus with specular reflectances of 0.0 and 1.0, respectively. For stimuli with a specular reflectance of 0.0, basic color terms beige and water were observed. The results were very similar to those observed in the categorical color-naming task (Experiment 1) within the ranges of colors used in Experiment 1. The frequent use of water is consistent with a previous report that showed that this term is commonly observed for Japanese subjects (Uchikawa & Boynton, 1987). For stimuli with a specular reflectance of 1.0, in addition to the ordinary basic color terms, GOLD, SILVER, and water were observed. Again, the results were similar to those observed in Experiment 1 for a specular reflectance of 1.0 except for the absence of COPPER. There was no other color term besides GOLD and SILVER specifically used for stimuli with a specular reflectance of 1.0 more than 50% of trials.

Table 1 shows the number of trials in which a given color term was used, number of subjects who used it, consistency, and maximum consensus for each color term. GOLD was used by all five subjects and its consistency (63%) and maximum consensus (90%) were comparable to those observed in Experiment 1. SILVER was also used by all five subjects and its consistency (86%) was higher than that of Experiment 1 (62%). This might be due to the

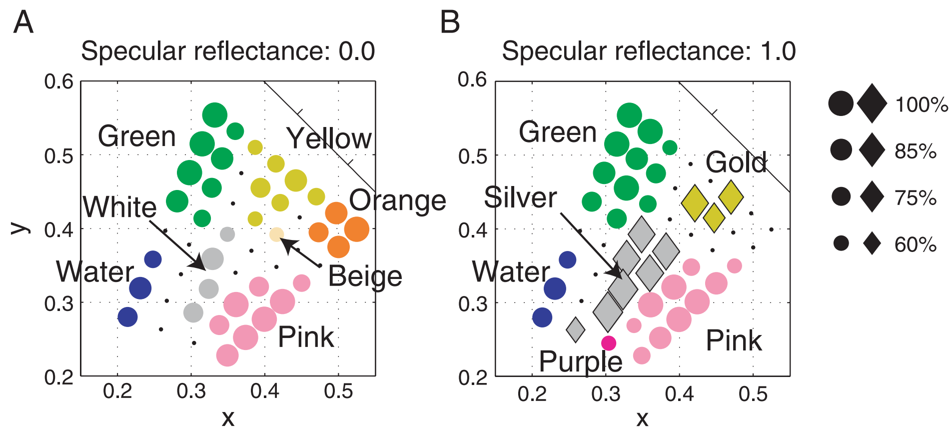

Figure 11. Results from the free color-naming task in Experiment 3 for stimuli with specular reflectances of (A) 0.0 and (B) 1.0 plotted on CIE *xy* chromaticity diagrams. Note that the scale on the chromaticity coordinates and the intervals of stimuli are different from those used in Experiment 1. The range of chromaticity coordinates examined is wider and the interval between each test point is larger in Experiment 3. The other conventions are the same as in Figure 2.

difference in the range of colors used: Experiment 3 contained more achromatic regions than that of Experiment 1 as can be seen in Figure 11. COPPER was used only by two subjects who participated in Experiment 1. Therefore, the use of COPPER might be biased by the instruction given in Experiment 1. Consistency and maximum consensus of basic color terms (yellow, green, pink, white, and orange) and beige were also comparable to those of Experiment 1. Among the other color terms used, water had high consistency (82%) and maximum consensus (90%) that was comparable to those of basic color terms, and this result is consistent with the observation in a previous study (Uchikawa & Boynton, 1987).

Discussion

In our daily life, we usually perceive the surface color of objects under a variety of illuminations and with different degrees of shading, glossiness, and texture. Previous studies have shown that, within natural scenes, humans can perceive surface colors by discounting these various surface features (Giesel & Gegenfurtner, 2010; Todd et al., 2004; Xiao & Brainard, 2008). By contrast, the color terms GOLD and SILVER are usually used only for glossy metallic surfaces (Beck, 1975), suggesting that humans do not simply ignore surface gloss to perceive surface colors. In this study, we used a categorical color-naming procedure to test whether GOLD and SILVER possess categorical properties known to be possessed by the basic color terms. The results showed that GOLD and SILVER were used for highly specular images with particular ranges of chromaticity and that they had categorical properties comparable to those of the basic color terms. This is the first finding that humans utilize surface gloss to categorize surface colors.

Using basic color terms to name matte surfaces

In Experiment 1, we conducted a color-naming task where the subjects were instructed to name the color of stimuli with different chromaticity coordinates and surface reflectances, by selecting a color term from among GOLD, SILVER, COPPER, 11 basic color terms, and beige. For the matte stimuli with a specular reflectance of 0.0 (Figure 2A), the naming was nearly the same as in a previous study in which uniform color plates were used as stimuli (Uchikawa & Boynton, 1987). We also conducted a color-naming task using uniform color patches with the same subjects (Figure 2B) and confirmed that names used for matte surfaces were very similar to those used for uniform color patches. Two measures of categorical properties, namely, consistency and consensus, were also similar to those observed in previous studies (Figure 5; Boynton & Olson, 1990; Uchikawa & Boynton, 1987), which indicates that 3D shape and shading did not affect categorical color naming. What is more, the naming did not change when the stimuli had a specular reflectance of 0.4 and contained highlights, which is consistent with an earlier report showing that the presence of highlights does not affect color matching between two images (Xiao & Brainard, 2008). Thus, when the specular reflectance is not so high, humans are able to discount the gloss to perceive the color of a surface.

Categorical properties of GOLD

In contrast to stimuli with low specular reflectance, when the specular reflectance was higher than 0.4, a substantial number of stimuli was named GOLD. For images with a specular reflectance of 1.0, GOLD was used for a large range of chromaticities centered at $x = 0.469$ and $y = 0.443$. A previous study similarly examined the ranges of chromaticity named GOLD using a color-naming procedure and found that GOLD was used for specific ranges of chromaticity with high specular reflectance (Nishizawa, 2007; Nishizawa et al., 2006). Our observations are consistent with those results.

In the present study, we examined the categorical properties of GOLD using three measures also employed in earlier studies: consistency, consensus, and reaction time (Boynton & Olson, 1990). Consistency is a measure of the stability of the use of color terms, consensus is a measure of how similarly color terms are used across subjects, and reaction time is a measure of the hesitation in the subjects' responses. With GOLD, all three of these measures were comparable to those of the basic color terms (Figures 5 and 6) and were higher than those for beige, a non-basic color term (Uchikawa & Boynton, 1987). This indicates that GOLD possesses categorical properties as strong as those possessed by the basic color terms. This is also confirmed by the free color-naming task examined in Experiment 3.

The dominant wavelength of GOLD was not identical to that of yellow or orange (Figure 4), and its color confusion pattern was also distinct from any of the basic color terms (Figure 7). This means that GOLD is not a subcategory of one of the basic color terms (e.g., glossy yellow); instead, it is a unique color category, distinct from any other color term. In addition, Experiment 2 showed that the rating of GOLD was not strongly affected by shape or illumination (Figure 10), which indicates that GOLD is associated with a particular category of surface properties, irrespective of an object's shape or illumination.

Several earlier studies have suggested that humans perceive surface colors by discounting various surface features caused by illumination and reflectance properties such as gloss (Giesel & Gegenfurtner, 2010; Todd et al., 2004; Xiao & Brainard, 2008). In the present study, we found that highly specular surfaces with a particular range of chromaticities could be named differently from matte surfaces. This finding suggests that humans do not simply discount surface gloss but can utilize gloss as a property of surfaces to categorize the colors of objects with particular colors and high specular reflectance. The results on the usage of GOLD and SILVER show that, in some cases, color categorization is tightly coupled to surface reflectance estimation.

Are GOLD and SILVER basic color terms?

The present results showed that the strength of categorical properties of GOLD and SILVER, as assessed by several psychophysical measures, was comparable to those of the basic color terms. However, whether GOLD and SILVER can be regarded as basic color terms must be carefully considered.

In contrast to ordinary basic color terms such as red or blue, GOLD and SILVER are also the name of an object, i.e., real gold and silver. As documented by Berlin and Kay (1969), the application of basic color terms should not be restricted to a narrow class of objects. Instead, basic color terms such as red or blue can be abstract and they should be able to apply to anything. Likewise, color terms GOLD and SILVER are not restricted to be used for the real gold/silver objects and used to describe colors of non-metallic objects such as golden hair or silver paper. In this regard, we cannot differentiate GOLD/SILVER from ordinary basic color terms. The situation seems similar to ORANGE, one of the ordinary basic color terms, that is a name of specific object, but it can also be applied to anything.

Another point is whether GOLD and SILVER are "necessary" to represent particular colors. Basic color terms are sufficient for partitioning color space (Boynton & Olson, 1990). This is true for ordinary basic color terms as long as uniform color plates were used. However, when the stimuli had high specular reflectance, subjects consistently employed other color terms, namely, GOLD and SILVER. One might argue that it might be possible for the subjects to name the color of stimuli using only the ordinary basic color terms even when the stimuli had high specular reflectance. However, the fact that the subjects consistently used GOLD and SILVER even in free color-naming experiment indicates that this is unlikely. Taken together, our present results are consistent with the idea that GOLD and SILVER share characteristics common with ordinary basic color terms.

One remaining issue is the cross-cultural nature of these terms. An important aspect of the basic color terms is their cross-cultural similarities (Berlin & Kay, 1969). In the present study, all of the participants were Japanese. Therefore, one question not answered by this study is whether there are cross-cultural differences in the use of metallic color terms, that is, to what degree are metallic color terms ubiquitous or variable across different cultures and languages.

Chromatic coordinates of GOLD category

The color term GOLD no doubt originated from the name of the material gold, but does the chromatic position of GOLD correspond to the chromaticity of real gold? To address this question, we plotted the naming data for GOLD obtained in Experiment 1 with the chromaticity coordinates of real gold calculated using the spectral reflectance of gold (Figure 12; see Appendix B for the calculation). The yellow diamonds represent the chromaticity coordinates named GOLD in more than 50% of trials in Experiment 1 when the specular reflectance was 1.0 (same as Figure 3C). The blue, green, and red dots are the calculated points for the chromaticity of real gold. Each color corresponds to different number of reflections (see Appendix B for the details). The chromaticity

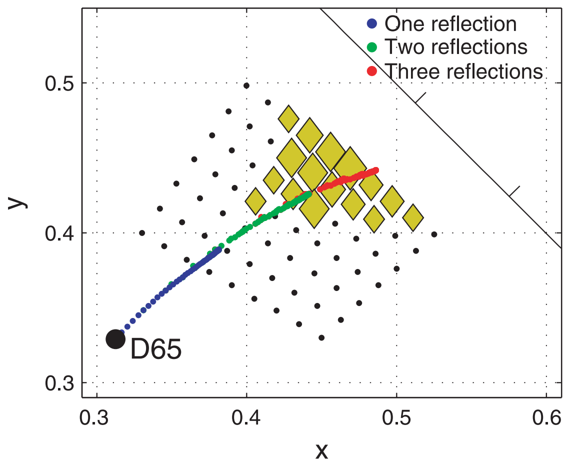

Figure 12. Comparison between the GOLD category observed in this study and the chromaticity of real gold materials. The yellow diamonds represent the chromaticities that were named GOLD in more than 50% of trials in Experiment 1 (same as Figure 3C). The blue, green, and red dots indicate the chromaticities of real gold calculated from the spectral reflectance of the metal. The chromaticity of light from a perfectly polished gold surface is indicated by the blue dots. The chromaticities of multireflected light from gold surfaces are indicated by the green or red dots, depending on the number of reflections (see Appendix B).

coordinates of real gold clearly overlap with the range of GOLD observed in this study. Presumably, this is because GOLD is closely related to the subjects' visual experience with golden objects.

On the other hand, we observed that the dominant wavelength of GOLD varied across the subjects and correlated with that of yellow and orange (Figure 4). This indicates that the GOLD category is not completely determined by experience; instead, a mechanism that determines the location of basic color terms may also induce the formation of the GOLD category.

Naming of the other metallic colors

SILVER and COPPER were also used for highly specular stimuli. The consistency, consensus, and reaction time for SILVER were comparable to those for GOLD and the basic color terms, so it can be said that SILVER also possesses categorical properties. On the other hand, the consistency and consensus for COPPER were lower and the reaction time was longer than those of the basic color terms, indicating that COPPER is not as clearly perceived categorically as the basic color terms. This may be related to the fact that objects made of copper are usually tarnished or oxidized, so we seldom encounter shiny copper objects. The naming of COPPER may be facilitated for images with textured patterns akin to oxidation, which would be in line with the previous finding that textural information facilitates the constant color perception of familiar objects (Hansen, Olkkonen, Walter, & Gegenfurtner, 2006; Olkkonen, Hansen, & Gegenfurtner, 2008).

Some basic color terms such as green and pink were also used for stimuli with high specular reflectance. There were no color terms specifically used for glossy surfaces in the range of colors named green or pink as examined in Experiment 3. In the natural environment, metals are the most common materials with high specular reflectance, but the colors of metals are restricted to the range of red, yellow, and white by the physical properties of metal, which make the reflectance of metals higher for longer wavelengths (Cox, 1987). In other words, metals in the natural environment rarely reflect only mid-wavelength (green) or short-wavelength (blue) light or absorb only mid-wavelength light (pink). The absence of color terms specifically used for green and pink glossy surfaces is likely related to this reality.

Relation between the categorical nature of GOLD and other perceptual attributes

This study examined the color naming of stimuli with different surface reflectances. To characterize the gloss, we used combinations of two parameters: diffuse and specular reflectance. In typical reflectance models, however, at least three parameters are used to describe surface gloss (diffuse reflectance, specular reflectance, and roughness in the Ward–Duer model used in this study). Examining the distributions of GOLD and SILVER within all three of these dimensions of reflectance space will further our understanding of the relationship between physical parameters and perceptual qualities (Nishizawa, 2007; Nishizawa et al., 2006). It would also be of interest to map GOLD and SILVER in a perceptual gloss space. A previous study defined a perceptually uniform 2D glossiness space by examining the dimensionality of perceived glossiness using the multidimensional scaling method (Ferwerda et al., 2001). By taking advantage of this space, it would be possible to compare categorical color naming and the perceptual uniformity of glossiness space. For the basic color terms, it has been shown that color category affects various cognitive tasks related to color vision, such as color discrimination (Özgen & Davies, 2002), color memory (Uchikawa & Shinoda, 1996), and visual search (Yokoi & Uchikawa, 2005). Whether these effects can also be observed with GOLD and SILVER remains an open question.

Our focus in this study was on GOLD, which can be defined based on surface gloss, though there are numerous other surface attributes, such as transparency (Fleming & Bülthoff, 2005; Motoyoshi, 2010) or texture (Ho, Landy, & Maloney, 2008), that can affect color naming. Although an increasing number of studies have focused on the perception of these complex surface attributes or the perception of materials (Adelson, 2001; Fleming & Bülthoff, 2005; Kim & Anderson, 2010; Motoyoshi, Nishida, Sharan, & Adelson, 2007; Nishida & Shinya, 1998; Wendt, Faul, Ekroll, & Mausfeld, 2010), the categorical nature of these attributes has gone unexplored, aside from a few recent studies (Sharan, Rosenholtz, & Adelson, 2009; Wolfe & Myers, 2010). If we can successfully manipulate the aforementioned surface attributes parametrically, it should be possible to examine the categorical properties not only of color terms but also material names using a naming procedure like that used in the present study.

Appendix A

Table A1.

			Number of named trials (max: 12)						
Specular	*x*	*y*	GOLD	COPPER	Green	Yellow	Brown	Orange	Mean RT (s)
0	0.4	0.498	0	0	9	3	0	0	1.923
	0.414	0.487	0	0	4	8	0	0	2.487
	0.428	0.476	0	0	3	9	0	0	2.75
	0.442	0.465	0	0	0	12	0	0	1.772
	0.456	0.454	0	0	0	10	0	2	1.581
	0.469	0.443	0	0	0	9	0	3	2.404
	0.483	0.432	0	0	0	3	0	9	2.543
	0.497	0.421	0	0	0	0	0	12	2.695
	0.511	0.41	0	0	0	0	0	12	2.293
	0.525	0.399	0	0	0	0	0	12	1.872
0.4	0.4	0.498	0	0	7	5	0	0	2.471
	0.414	0.487	0	0	5	7	0	0	2.748
	0.428	0.476	0	0	1	11	0	0	2.095
	0.442	0.465	0	0	0	11	0	1	1.939
	0.456	0.454	0	0	0	10	0	2	2.441
	0.469	0.443	0	0	0	10	1	1	4.239
	0.483	0.432	0	0	0	6	0	6	3.662
	0.497	0.421	1	0	0	2	0	9	3.03
	0.511	0.41	0	0	0	1	0	11	2.061
	0.525	0.399	1	0	0	0	0	11	2.567
0.8	0.4	0.498	0	0	10	2	0	0	1.972
	0.414	0.487	0	0	3	9	0	0	3.333
	0.428	0.476	2	0	0	10	0	0	2.193
	0.442	0.465	2	0	0	10	0	0	2.558
	0.456	0.454	5	0	0	5	0	2	3.403
	0.469	0.443	7	0	0	3	0	2	3.477
	0.483	0.432	7	0	0	1	0	4	3.029
	0.497	0.421	6	0	0	0	0	6	2.269
	0.511	0.41	6	1	0	0	0	5	2.075
	0.525	0.399	3	1	0	0	0	8	3.348
1	0.4	0.498	0	0	10	2	0	0	2.175
	0.414	0.487	2	0	6	4	0	0	2.616
	0.428	0.476	7	0	2	3	0	0	3.989
	0.442	0.465	9	0	0	3	0	0	2.424
	0.456	0.454	10	0	0	2	0	0	2.692
	0.469	0.443	11	0	0	0	0	1	2.013
	0.483	0.432	8	1	0	0	2	1	2.509
	0.497	0.421	8	0	0	0	1	3	2.476
	0.511	0.41	7	1	0	0	0	4	3.087
	0.525	0.399	6	1	0	0	1	4	3.803

Table A1. The number of times a categorical color name was applied to each color and the mean reaction time for each stimulus in Experiment 1. Only the results for saturated colors are shown because the number of stimuli is large.

Appendix B

According to Fresnel equation, reflectance spectrum of metals depends on the incident angle of lights. We calculated the reflectance spectrum with varying angles of incident lights (0–90 deg with 1 deg step) using the refractive indexes listed on a reference (Johnson & Christy, 1972). Then, the chromaticity coordinates of the reflected light for each angle of the incident light was calculated assuming that the gold surface was under CIE Standard Illuminants D65 with no polarization.

It should be noted, however, that the chromaticity coordinates of the light reflected from an object surface depends on how many times the light is reflected on the surface. When a gold surface is perfectly smooth, the illuminating light will be reflected only once by the surface before being captured by the observer's eyes (one-reflection light). Consequently, the spectral distribution of one-reflection light will be a simple convolution of the illumination's spectral distribution and the surface reflectance spectrum. The chromaticity coordinates of one-reflection light for each angle of the incident light are plotted as the blue dots in Figure 12. If a gold surface is not smooth, the illumination light could be reflected more than once due to the geometrical structure of the surface. When the light is reflected twice at the surface before being captured by the observer's eyes (two-reflection light), the spectral distribution of the light is the convolution of the one-reflection light with the surface reflectance spectrum (the green dots in Figure 12). We assumed that the angles of the incident lights for the first and second reflections have no correlation and calculated the chromaticities for all combinations of the two angles. In the same manner, we can calculate the coordinates of three-reflection light (the red dots) and so on. The chromaticity coordinates of the reflected lights from the gold surfaces will be located around the ensemble of these dots depending on the incident angle of lights and the fine structure as well as the geometry of the gold surfaces.

Acknowledgments

We thank Dr. Hiromi Okamoto for comments on the calculation of spectral reflectances of metals. This work was supported by a Grant-in-Aid for Scientific Research on Innovative Areas (No. 22135007) from the Ministry of Education, Science, Culture, Sports and Science, Japan.

Commercial relationships: none.
Corresponding author: Hidehiko Komatsu.
Email: komatsu@nips.ac.jp.
Address: Division of Sensory and Cognitive Information, National Institute for Physiological Sciences, Myodaiji, Okazaki-shi, Aichi 444-8585, Japan.

References

Adelson, E. H. (2001). On seeing stuff: The perception of materials by humans and machines. In B. E. Rogowitz & T. N. Pappas (Eds.), *Proceedings of the SPIE. Volume 4299: Human Vision and Electronic Imaging VI* (pp. 1–12). Bellingham, WA: SPIE.

Ayama, M., Nakatsue, T., & Kaiser, P. K. (1987). Constant hue loci of unique and binary balanced hues at 10, 100, and 1000 Td. *Journal of the Optical Society of America A, 4,* 1136–1144. [PubMed]

Beck, J. (1975). The perception of surface color. *Scientific American, 1975,* 10.

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

Boynton, R. M., MacLaury, R. E., & Uchikawa, K. (1989). Centroids of color categories compared by two methods. *Color Research & Application, 14,* 6–15.

Boynton, R. M., & Olson, C. X. (1987). Locating basic colors in the OSA space. *Color Research & Application, 12,* 94–105.

Boynton, R. M., & Olson, C. X. (1990). Salience of chromatic basic color terms confirmed by three measures. *Vision Research, 30,* 1311–1317. [PubMed]

Cox, P. A. (1987). *The electronic structure and chemistry of solids*. New York: Oxford University Press.

Debevec, P. E. (1998). Rendering synthetic objects into real scenes: Bridging traditional and image-based graphics with global illumination and high dynamic range photography. *Proceedings of SIGGRAPH, 1998,* 189–198.

Ferwerda, J. A., Pellacini, F., & Greenberg, D. P. (2001). A psychophysically-based model of surface gloss perception. *Proceedings of SPIE Human Vision and Electronic Imaging, 4299,* 291–301.

Fleming, R. W., & Bülthoff, H. H. (2005). Low-level image cues in the perception of translucent materials. *ACM Transactions on Applied Perception, 2,* 346–382.

Giesel, M., & Gegenfurtner, K. R. (2010). Color appearance of real objects varying in material, hue, and shape. *Journal of Vision, 10*(9):10, 1–21, http://www.journalofvision.org/content/10/9/10, doi:10.1167/10.9.10. [PubMed] [Article]

Guest, S., & Van Laar, D. (2000). The structure of colour naming space. *Vision Research, 40,* 723–734. [PubMed]

Hansen, T., Olkkonen, M., Walter, S., & Gegenfurtner, K. R. (2006). Memory modulates color appearance. *Nature Neuroscience, 9,* 1367–1368. [PubMed]

Hansen, T., Walter, S., & Gegenfurtner, K. R. (2007). Effects of spatial and temporal context on color categories and color constancy. *Journal of Vision, 7*(4):2, 1–15, http://www.journalofvision.org/content/7/4/2, doi:10.1167/7.4.2. [PubMed] [Article]

Ho, Y. X., Landy, M. S., & Maloney, L. T. (2008). Conjoint measurement of gloss and surface texture. *Psychological Science, 19,* 196–204. [PubMed]

Johnson, P. B., & Christy, R. W. (1972). Optical constants of the noble metals. *Physical Review B, 6,* 4370–4379.

Kim, J., & Anderson, B. (2010). Image statistics and the perception of surface gloss and lightness. *Journal of Vision, 10*(9):3, 1–17, http://www.journalofvision.org/content/10/9/3, doi:10.1167/10.9.3. [PubMed] [Article]

Motoyoshi, I. (2010). Highlight–shading relationship as a cue for the perception of translucent and transparent materials. *Journal of Vision, 10*(9):6, 1–11, http://www.journalofvision.org/content/10/9/6, doi:10.1167/10.9.6. [PubMed] [Article]

Motoyoshi, I., Nishida, S., Sharan, L., & Adelson, E. H. (2007). Image statistics and the perception of surface qualities. *Nature, 447,* 206–209. [PubMed]

Motoyoshi, I., Nishizawa, T., & Uchikawa, K. (2007). Specular reflectance and the perception of metallic surfaces [Abstract]. *Journal of Vision, 7*(9):451, 451a, http://www.journalofvision.org/content/7/9/451, doi:10.1167/7.9.451.

Nishida, S., & Shinya, M. (1998). Use of image-based information in judgments of surface-reflectance properties. *Journal of the Optical Society of America A, Optics, Image Science, and Vision, 15,* 2951–2965. [PubMed]

Nishizawa, T. (2007). *Mechanisms of metal perception* [in Japanese]. Master thesis, Tokyo Institute of Technology.

Nishizawa, T., Segawa, K., & Uchikawa, K. (2006). Metric for gold perception. *VISION, 18,* Supplement, 134.

Olkkonen, M., Hansen, T., & Gegenfurtner, K. R. (2008). Color appearance of familiar objects: Effects of object shape, texture, and illumination changes. *Journal of Vision, 8*(5):13, 1–16, http://www.journalofvision.org/content/8/5/13, doi:10.1167/8.5.13. [PubMed] [Article]

Olkkonen, M., Hansen, T., & Gegenfurtner, K. R. (2009). Categorical color constancy for simulated surfaces. *Journal of Vision, 9*(12):6, 1–18, http://www.journalofvision.org/content/9/12/6, doi:10.1167/9.12.6. [PubMed] [Article]

Olkkonen, M., Witzel, C., Hansen, T., & Gegenfurtner, K. R. (2010). Categorical color constancy for real surfaces. *Journal of Vision, 10*(9):16, 1–22, http://www.journalofvision.org/content/10/9/16, doi:10.1167/10.9.16. [PubMed] [Article]

Özgen, E., & Davies, I. R. L. (2002). Acquisition of categorical color perception: A perceptual learning approach to the linguistic relativity hypothesis. *Journal of Experimental Psychology, 131,* 477–493. [PubMed]

Sharan, L., Rosenholtz, R., & Adelson, E. H. (2009). Material perception: What can you see in a brief glance? [Abstract]. *Journal of Vision, 9*(8):784, 784a, http://www.journalofvision.org/content/9/8/784, doi:10.1167/9.8.784.

Todd, J. T., Norman, J. F., & Mingolla, E. (2004). Lightness constancy in the presence of specular highlights. *Psychological Science, 15,* 33–39. [PubMed]

Troost, J. M., & de Weert, C. M. (1991). Naming versus matching in color constancy. *Perception & Psychophysics, 50,* 591–602. [PubMed]

Uchikawa, H., Uchikawa, K., & Boynton, R. M. (1989). Influence of achromatic surrounds on categorical perception of surface colors. *Vision Research, 29,* 881–890. [PubMed]

Uchikawa, K., & Boynton, R. M. (1987). Categorical color perception of Japanese observers: Comparison with that of Americans. *Vision Research, 27,* 1825–1833. [PubMed]

Uchikawa, K., & Shinoda, H. (1996). Influence of basic color categories on color memory discrimination. *Color Research & Application, 21,* 430–439.

Ward, G. J. (1994). The RADIANCE lighting simulation and rendering system. *Computer Graphics, 28,* 459–472.

Wendt, G., Faul, F., Ekroll, V., & Mausfeld, R. (2010). Disparity, motion, and color information improve gloss constancy performance. *Journal of Vision, 10*(9):7, 1–17, http://www.journalofvision.org/content/10/9/7, doi:10.1167/10.9.7. [PubMed] [Article]

Wolfe, J. M., & Myers, L. (2010). Fur in the midst of the waters: Visual search for material type is inefficient. *Journal of Vision, 10*(9):8, 1–9, http://www.journalofvision.org/content/10/9/8, doi:10.1167/10.9.8. [PubMed] [Article]

Xiao, B., & Brainard, D. H. (2008). Surface gloss and color perception of 3D objects. *Visual Neuroscience, 25,* 371–385. [PubMed]

Yokoi, K., & Uchikawa, K. (2005). Color category influences heterogeneous visual search for color. *Journal of the Optical Society of America A, Optics, Image Science, and Vision, 22,* 2309–2317. [PubMed]