Despite numerous prior studies, important questions about the Japanese color lexicon persist, particularly about the number of Japanese basic color terms and their deployment across color space. Here, 57 native Japanese speakers provided monolexemic terms for 320 chromatic and 10 achromatic Munsell color samples. Through $k$-means cluster analysis we revealed 16 statistically distinct Japanese chromatic categories. These included eight chromatic basic color terms ($\text{aka}/\text{red}$, $\text{ki}/\text{yellow}$, $\text{midori}/\text{green}$, $\text{ao}/\text{blue}$, $\text{pink}$, $\text{orange}$, $\text{cha}/\text{brown}$, and $\text{murasaki}/\text{purple}$) plus eight additional terms: $\text{mizu}$ (”water”), $\text{hada}$ (”skin tone”), $\text{peach}$, $\text{kon}$ (”indigo”), $\text{kon}$ (”indigo”), $\text{teal}$, $\text{lavender}$, $\text{magenta}$, $\text{lime}$) that had no equivalent in the other. Our analysis revealed two statistically distinct Japanese motifs (or color-naming systems), which differed mainly in the extension of $\text{mizu}$ across our color palette. Comparison of the present data with an earlier study by Uchikawa & Boynton (1987) suggests that some changes in the Japanese color lexicon have occurred over the last 30 years.

**Introduction**

The association between color terms, the color categories they name, and the stimuli that elicit them is a color term. Japanese and American English color-naming systems were broadly similar, except for color categories in one language ($\text{mizu}$, $\text{kon}$, $\text{teal}$, $\text{lavender}$, $\text{magenta}$, $\text{lime}$) that had no equivalent in the other. Our analysis revealed two statistically distinct Japanese motifs (or color-naming systems), which differed mainly in the extension of $\text{mizu}$ across our color palette. Comparison of the present data with an earlier study by Uchikawa & Boynton (1987) suggests that some changes in the Japanese color lexicon have occurred over the last 30 years.
classic model system for studying the relationship between words and their referents. This is because all languages have at least some color terms in their lexicons, because colors are easily specified quantitatively, and because the physiology of the perception of color is better understood than the perception of many other stimuli. Furthermore, a physiological response to color categories may be present even in prelinguistic infants (Yang, Kanazawa, Yamaguchi, & Kuriki, 2016), although there remains some controversy about whether the acquisition of language modifies those innate categories (Franklin, Clifford, Williamson, & Davies, 2005; Roberson, Davidoff, Davies, & Shapiro, 2006).

Much of the modern work on color and language has been inspired by three key proposals in the seminal work of Berlin and Kay (1969). The first is that world languages contain salient terms for colors—the basic color terms (BCTs)—that are monolexemic, that are known and used by all members of the language community, that can be used to communicate about the color of any type of object, and that name colors not covered by any other BCT. The second proposal is that the BCTs in every language name colors that are derived from a set of 11 universal categories. The third proposal is that cross-cultural differences in color naming exist because color lexicons are at different stages along a constrained trajectory of color-term evolution. As color lexicons evolve over time, they increase in size, adding BCTs in a highly constrained order. With this “universalist” framework in mind, we have examined the modern Japanese color lexicon. We compare it to the contemporaneous American English color lexicon, and we compare it to an earlier study of the Japanese color lexicon for evidence of recent evolution of Japanese color terms.

The study of the Japanese color lexicon is important for three interrelated reasons. First, Japanese is spoken in a modern, highly industrialized society, where the chromatic environment is as diverse and colorful as anywhere on earth. According to the universalist perspective, the Japanese color lexicon should therefore closely approximate the lexicons of English and other languages spoken in industrialized societies. Second, there remain several questions regarding the number of BCTs in the Japanese color lexicon. It is known from the earliest written records of vernacular Japanese (the Manyō-shū poems, dating from before 759 D.E.) that the Japanese words ao (blue) and midori (green) were used more or less interchangeably, in a usage pattern similar to the “grue” motif of Lindsey and Brown (2009) for color-naming systems with a single term for green-or-blue. In present-day Japanese, ao is still used to denote certain green things, as well as being an abstract color term for blue things in general, whereas midori always names only green things. Moreover, historical linguists (e.g., Stanlaw, 2010) sometimes include kon (indigo) as a word for dark blue among Japanese BCTs. Therefore, it is possible that Japanese, like many world languages spoken in nonindustrialized societies, might not conform to the color-category structure seen in English. Third, a quantitative, empirical study of Japanese color naming that was conducted 30 years ago by Uchikawa and Boynton (1987; U&B) suggested that three nonbasic color terms—mizu (light blue), hada (peach), and kusa (yellow green)—might achieve BCT status sometime in the future. Comparing the results of the present study to those of U&B allows us to examine the Japanese color lexicon for evidence of language change over the intervening years.

U&B investigated Japanese color naming from the universalist perspective of Berlin and Kay. Using a color palette consisting of the 425 samples comprising the OSA-UCS (Optical Society of America, Uniform Color Scale), U&B found that Japanese color terms conforming to Berlin and Kay’s 11 BCTs showed better across-subjects consensus, better test–retest reliability, and shorter reaction time than other nonbasic Japanese color terms, including mizu, hada, kusa, and kon. The present study diverges from U&B’s methodology in two important respects. First, whereas U&B’s samples spanned a relatively narrow range of generally low chromas (saturations), the present study adopts the 330 Munsell samples that were used by Lindsey and Brown (2014; L&B) and most other modern studies of color naming, which contain a larger range of generally higher chromas.

Second, the present study applies several quantitative tools that Lindsey and Brown (2006, 2009, 2014) have developed over the years for analyzing color-naming data. They showed that cluster analysis and associated statistical techniques can reveal important regularities in a language’s color lexicon that might be missed by ethnographic studies or analyses of frequency of word usage alone (Lindsey & Brown, 2006). In particular, cluster analysis offers an objective way of controlling for synonymy in color naming by ignoring the color terms themselves and focusing instead on how they are deployed across color space. In the languages that Lindsey and Brown have examined so far—the 110 languages spoken in nonindustrialized societies that were included in the World Color Survey (Kay, Berlin, Maffi, Merrifield, & Cook, 2009 [WCS]; Lindsey & Brown, 2006) as well as English (L&B), Somali (Brown, Isse, & Lindsey, 2016), and Hadzane (Lindsey, Brown, Brainard, & Apicella, 2015)—the color terms glossed by cluster analysis correspond, with minor variations and some additions, to the standard list of universal BCTs from Berlin and Kay.

Lindsey and Brown (2009) applied a second cluster analysis to the color lexicons of the WCS, revealing the existence of a limited number of common color-naming systems, which they called motifs. These motifs recur,
with minor variation, throughout the WCS data set. Strikingly, almost all the languages in the WCS, as well as Somali (Brown et al., 2016) and English (L&B), contain multiple motifs among their speakers. The importance of the motif analysis is that it can reveal statistically significant regularities in subpopulations of informants in a diverse language community that would be missed if that community were assumed to be homogeneous. In the case of American English, speakers’ color vocabularies are divided into two motifs. Those motifs refine Berlin and Kay’s concept of the BCT in that, in addition to the 11 original BCTs, some color terms (e.g., teal, lavender, peach, and maroon) are “basic” for the individuals whose color idiolects fall into one motif but not for those whose color idiolects fall into the other motif.

The diversity of color idiolects seen in American English and elsewhere, as embodied in the motif concept, suggests a mechanism for color-term evolution that parallels biological evolution: Color lexicons evolve when the proportion of speakers in motifs with fewer BCTs declines and the proportion of speakers in motifs with more BCTs increases. It is from this theoretical perspective that we compare the present structure of the Japanese color lexicon to that observed 30 years ago by U&B.

### Methods

#### Subjects

Thirty-two subjects (18 men, 14 women) from Tohoku University and 25 from the Tokyo Institute of Technology (12 men, 13 women) took part in this study. All were native speakers of Japanese. Most subjects were graduate students, but Japanese authors IK, YM, KF, and RT also participated. Only the authors were aware of the purposes of the study at the time they were tested. All subjects had normal or corrected-to-normal visual acuity, and their color vision was confirmed to be normal with Ishihara pseudoisochromatic plates. The experimental procedures followed the precepts of the Declaration of Helsinki and were approved by the ethics committees of Tohoku University and the Tokyo Institute of Technology.

#### Apparatus

**Color samples and illuminant**

The color samples, illuminant, and background color papers were similar to those used in the WCS. The 330 color chips used in the present study were from the Munsell Book of Color glossy ed., X-Rite, Inc., www.munsell.com). The chips were chosen to match the WCS samples with respect to hue, chroma, and value (although the WCS samples were from the matte edition). Each chip was mounted on a cardboard square 5 cm by 5 cm covered with gray matte paper approximating N5/ (in Munsell notation). Four pairs of 40-W D65-simulating fluorescent lamps with high color-rendering index (FLR40S-D-EDL-D65, Toshiba, Minato, Japan) were mounted on the ceiling of an observation booth, providing an illuminance of 2,713 lx. An amber filter covered four lamps to adjust the color temperature of the illuminant to approximate 6000 K.

#### Procedure

Subjects used a single, monolexemic color term to name each sample. They were not allowed to use compound color terms like ki-midori (yellow-green) or modifier words like usu-murasaki (pale purple). However, they were allowed to use the name of a substance if they felt that the name was generally agreed to represent a color and could be generalized to name the color of any type of object. Each session took about 40 min.

#### Cluster analysis

Analysis of the Japanese color-naming data was performed in two steps, both of which involved $k$-means cluster analysis. The first cluster analysis was used to extract two entities from the raw data sets: (a) an estimate of the number of statistically significant named chromatic color categories in the Japanese language and (b) the extensions of each of these categories across color space. The second step used the results of the first cluster analysis to examine the color-naming patterns (motifs) used by Japanese informants by (a) estimating the number of motifs and (b) determining their categorical structures.

All analyses were performed using MATLAB (MathWorks, Natick, MA) and Mathematica (Wolfram Research, Inc., Champaign, IL) software platforms. We used custom programs which had been used previously by Lindsey and Brown and their colleagues in their analyses of color-naming data (Brown et al., 2016; Lindsey & Brown, 2006, 2009, 2014). Here we present an overview of our methodology; additional details may be found in the specific references cited. The first $k$-means cluster analysis was used to classify feature vectors representing the sets of color samples associated with each chromatic color term deployed by each of our Japanese informants. A chromatic color
term was defined as a term used by a subject to name one or more of the 320 chromatic colors in the WCS chart but never used by that subject to name any of the 10 achromatic colors (achromatic color terms were handled separately, as described later). Each chromatic-term feature vector consisted of 320 elements, each of which was set to a value of 1 or 0 depending on whether (or not) the chromatic color term was used by the informant to name the WCS color sample represented by that particular vector element (for details, see Lindsey & Brown, 2006, 2009, 2014). The resulting 828 binary feature vectors obtained from the chromatic words used by our 57 Japanese subjects were then sorted into $k$ clusters using the kmeans() function in MATLAB.

This first $k$-means cluster analysis was designed to control for synonymy and homonymy in estimating the number of statistically significant named chromatic color categories in the Japanese language, which we designate $k_{L_opt}$. Cluster analysis classifies responses solely on the basis of how color terms are deployed across the 320 WCS chromatic colors, as embodied in the patterns of color-term deployment encoded in the binary feature vectors, without regard for the actual terms used by the subject. In American English (L&B), for example, $k$-means analysis showed that cyan and turquoise were synonymous with teal. The same analysis also revealed that tan has two meanings in American English; some subjects used it to name greenish-brown colors, which $k$-means analysis assigned to the olive English color category. Other subjects used tan to name light, pale pinkish-orange colors, and these feature vectors were assigned to the beige category. As we show later, the Japanese color lexicon also contains chromatic synonyms and homonyms.

By design, the $k$-means algorithm will produce a cluster solution for any predetermined number of clusters $k$ from 1 to the total number $N$ of feature vectors being sorted. Thus, $k$-means cannot estimate $k_{L_opt}$ without additional analysis. For this purpose, we relied on the gap statistic of Tibshirani, Walther, Hastie (2001). First we performed $k$-means analyses for values of $k$ from 1 to 25. Then, following the computational framework of Tibshirani et al., we performed gap-statistic analysis on these 25 separate cluster results by comparing, for each value of $k$, the tightness of clustering of the data to the tightness obtained by $k$-means clustering (using the same value of $k$) of reference null distributions derived from the data, as described later. By design, the expected value of $k_{L_opt}$ for a reference distribution is 1. Thus, as the value of $k$ increases from 1 to $k_{L_opt}$, the tightness of clustering of the data is expected to improve relative to that obtained from $k$-means clustering of the reference null distributions. Beyond $k_{L_opt}$, increasing $k$ should not lead to any further improvement in the relative tightness of clustering. We express this result with the gap statistic $G(k)$ (see L&B, equation 2): $G(k) \geq 0.0$, $2 \leq k \leq k_{L_opt}$. A step-by-step computational framework for gap-statistic analysis is given by Tibshirani et al. (2001, pp. 414–415). See L&B (pp. 11–14) for additional details regarding our particular implementation of $k$-means/gap-statistic analysis.

A somewhat subjective aspect of our methodology involved the algorithm for creating suitable reference null distributions for the gap-statistic analysis. For the present study, we adopted the algorithm used by Lindsey and Brown (2006, 2009, 2014), which is similar to one used by Kay and Regier (2003) in their analysis of color-naming centroids obtained from the data of WCS informants. To create a reference null distribution, each informant’s raw chromatic color terms were first arranged in a $40 \times 8$ matrix according to the informant’s responses to the 40 Munsell hues $\times 8$ Munsell lightnesses of the chromatic samples used in our study. We then circularly shifted the elements of each informant’s matrix by a random number of columns on the ”cylindrical” surface of the WCS color space (shifting in the hue dimension) and then randomly reflected the rows of the resulting matrix (flipped the matrix vertically, corresponding to the lightness dimension) either zero times or one time. The resulting matrix was then decomposed into the appropriate feature vectors, as outlined earlier, based on the new mapping of color terms onto the WCS colors. In this way, our reference null distributions preserved much of the basic structure of patterns of Japanese color-term deployment—for example, the sizes and shapes of the patterns—while randomizing their locations within the 2-D hue/lightness coordinate frame of the WCS color chart. In this way, we obtained reference distributions with expected numbers of clusters of 1.

L&B noted some variation in the solutions produced from run to run in their $k$-means/gap-statistic analysis of the American English lexicon. Therefore, following their approach, we performed 1,000 different $k$-means/gap-statistic analyses, as described earlier (see Figure 4). The results of this procedure were compiled into a histogram of 1,000 resulting estimates of $k_{L_opt}$. Our conclusions regarding the size and structure of the Japanese color lexicon were based on the modal value of this histogram (Figure 4, inset). As we show later, the first step in our analysis of Japanese color naming revealed 16 distinct clusters of chromatic color-term deployment.

In this approach, the color samples associated with the feature vectors assigned to each of the 16 clusters define the extensions of 16 chromatic color categories in color space. These collections of color samples can also be compiled and displayed as consensus plots (e.g.,
Figure 5) representing the color categories. Moreover, the feature-vector clusters can be given names, and thus may be used as a glossary for evaluating the diversity of color terms deployed by informants. We chose common Japanese color terms as names for each of the clusters: aka (red), ao (blue), ki (yellow), and so on (see later for details). The samples contained in the clusters and the most frequent color terms associated with them thus form a glossary for classifying, within the context of our cluster analysis, all the different terms used by Japanese subjects. For example, if the feature vector for the color term moegi, as deployed by a particular informant, falls into the midori (green) cluster, then we say that “moegi glosses to midori.”

To obtain the full Japanese glossary, we added three achromatic categories to the 16 chromatic categories determined by cluster analysis. The achromatic categories were defined a priori in the following way: The Japanese white category was defined as the set of colors that included the lightest among the achromatic WCS colors. Black was defined as the set of colors that included the darkest achromatic sample, and gray was defined as the set of color samples that included one or more of the remaining eight achromatic samples that were neither white nor black.

We next performed a second k-means/gap-statistic analysis to determine $k_{M,\text{opt}}$, the number of statistically significant motifs in the Japanese color lexicon, and the structures of these motifs. Our conclusions were based on clustering 57 motif feature vectors, each vector representing all 330 color-naming responses of a single Japanese subject. Each feature vector comprised 19 elements corresponding to the 19 color categories derived from the first step of our analysis: the 16 glossed chromatic categories (derived from the first cluster analysis) plus the three defined achromatic color categories. Each of the 19 elements was assigned a value between 0.0 and 1.0, which was the proportion of samples (out of 330 WCS samples) a given subject named with the glossed color term. For example, if the subject used the word aka to name three samples, the value of the aka element in that subject’s motif feature vector would receive the number $3/330 = 0.0091$. The 57 feature vectors were then sorted into $k$ clusters using the kmeans(.) MATLAB function. Gap-statistic analysis was used to determine $k_{M,\text{opt}}$ based on k-means results obtained for $k = 1, \ldots, 5$.

To determine $k_{M,\text{opt}}$ for the second step of our analysis, the gap-statistic analysis was based on motif reference null distributions created by randomization of each of the 57 subject motif feature vectors. Randomization was accomplished by random permutation (scrambling the order) of the 19 values in each subject’s motif feature vector. Our final estimate of the motifs in the Japanese color lexicon was based on the modal result from 1,000 separate k-means/gap-statistic analyses. Further details are given by L&B (pp. 14–17). We obtained identical results from each of these 1,000 analyses. Subjects’ glossed color-naming patterns were then compiled within motifs, and aggregate results were displayed as consensus diagrams (e.g., Figure 6).

### Results

#### Descriptive statistics

The 57 Japanese subjects used a total of 93 unique color terms. L&B’s 51 English-speaking subjects used 122 unique terms. We compared the two studies based on equal numbers of subjects by calculating the average number of color terms (and 95% confidence interval) from 10,000 random $N = 51$ samples of our 57 Japanese subjects. The simulation yielded a value of 88.6 (95% CI [67, 93]) unique words per 51 Japanese subjects. In an effort to control for idiosyncratic responses, L&B also determined that 43 terms were used by four or more of the 51 subjects. In the present study, 32 terms were used by four or more of the 57 Japanese subjects, with an average of 30.4 (95% CI [28, 32]) terms per 51 subjects. U&B’s 10 Japanese subjects used a total of 66 unique terms, a value that falls within the range of our simulations (10,000 random $N = 51$ samples of 57 Japanese subjects), which yielded an average of 43 (95% CI [25, 72]) terms per 10 subjects. Thus, Japanese subjects in the present study used slightly fewer terms than were used by L&B’s American subjects, but their color-term usage falls in line with that of U&B’s Japanese subjects.

A histogram of color-term usage by four or more subjects in our study is shown in Figure 1. Table 1 shows the number of subjects who used each color term (words in parentheses show direct English translations) and the number of samples receiving each color term. Japanese equivalents of all of Berlin and Kay’s (1969) BCTs were used by all 57 informants, with the exception of orange, pink, kuro (black), and hai (gray). Orange ($n = 56$) and pink ($n = 55$) have all but replaced the traditional Japanese terms daidai and momo, respectively. Kuro was used by 53 subjects, the remaining four subjects preferring instead terms meaning gray: nezumi, hai, and/or gray. Hai was used by 52 subjects; the remaining five subjects used gray instead. Mizu (light blue; $n = 55$) and to a lesser extent hada (skin tone; $n = 48$) complete the inventory of terms used by a large majority of our Japanese subjects.

All subjects in the present study used 11 or more color terms (Figure 2). The mean number of unique terms per subject was $17.68 \pm 6.7$ (mode = 16). This number was smaller than the average 21.9 (mode = 19) color terms found in L&B’s study of American English speakers.
Rank-order analysis

Many investigators have observed that frequency of word usage often obeys a power law: It falls on a straight line in log-log coordinates when plotted as a function of the rank order of word usage (Mitzenmacher, 2003). Figure 3 shows a log-log plot of the number of subjects who used each color term (the popularity of the color term) versus the rank order of that term’s popularity. The flat part on the top, plotted with triangles, represents color terms that were used with very high consensus, while the steeply sloped distribution plotted with circles represents color terms used with lower consensus. Twelve color terms—the 11 BCTs plus mizu (light blue)—were used by >89% of informants. On the falling limb, there is a gap between hada (skin tone) and oudo (light brown), and a larger gap between oudo and the next most popular term, kon (dark blue). U&B discussed mizu and hada as possible BCTs, and we consider them further later. U&B also suggested kusa (yellow-green) as another candidate BCT, but kusa was used by few informants (n = 3) in the present study.

This pattern of color-term popularity in Figure 3 differs slightly from what L&B found for the American English lexicon. Their raw data were best fitted with three straight-line segments, which suggested three domains of American English color-term usage: a horizontal segment corresponding to highly popular BCTs, a modestly negatively sloping line segment corresponding to popular but emerging BCTs, and a steeply negatively sloping segment representing low-popularity nonbasic color terms. In contrast, the present data were fitted with only two segments, suggesting a set of established BCTs (the 11 BCTs of B&K plus mizu), followed by a steeply declining

<table>
<thead>
<tr>
<th>Color term (common meaning)</th>
<th>Subjects</th>
<th>Total samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>shiro</td>
<td>57</td>
<td>1,020</td>
</tr>
<tr>
<td>aka</td>
<td>57</td>
<td>723</td>
</tr>
<tr>
<td>ki</td>
<td>57</td>
<td>873</td>
</tr>
<tr>
<td>midori</td>
<td>57</td>
<td>4,415</td>
</tr>
<tr>
<td>ao</td>
<td>57</td>
<td>1,844</td>
</tr>
<tr>
<td>cha (fruit of the tea plant)</td>
<td>57</td>
<td>1,257</td>
</tr>
<tr>
<td>murasaki</td>
<td>57</td>
<td>2,389</td>
</tr>
<tr>
<td>orange</td>
<td>56</td>
<td>792</td>
</tr>
<tr>
<td>mizu (water)</td>
<td>55</td>
<td>1,367</td>
</tr>
<tr>
<td>pink</td>
<td>55</td>
<td>1,548</td>
</tr>
<tr>
<td>kuro</td>
<td>53</td>
<td>422</td>
</tr>
<tr>
<td>hai (ashes)</td>
<td>52</td>
<td>268</td>
</tr>
<tr>
<td>hada (skin)</td>
<td>48</td>
<td>413</td>
</tr>
<tr>
<td>oudo (sand/mud)</td>
<td>36</td>
<td>210</td>
</tr>
<tr>
<td>kon (indigo)</td>
<td>21</td>
<td>84</td>
</tr>
<tr>
<td>cream</td>
<td>17</td>
<td>72</td>
</tr>
<tr>
<td>gray</td>
<td>15</td>
<td>54</td>
</tr>
<tr>
<td>matcha (powdered ceremonial tea)</td>
<td>12</td>
<td>92</td>
</tr>
<tr>
<td>yamabuki (globeflower)</td>
<td>12</td>
<td>38</td>
</tr>
<tr>
<td>enji (carmine/cochineal)</td>
<td>11</td>
<td>38</td>
</tr>
<tr>
<td>shu (vermilion)</td>
<td>11</td>
<td>27</td>
</tr>
<tr>
<td>daidai (bitter orange)</td>
<td>8</td>
<td>53</td>
</tr>
<tr>
<td>beige</td>
<td>8</td>
<td>12</td>
</tr>
<tr>
<td>sora (sky)</td>
<td>8</td>
<td>81</td>
</tr>
<tr>
<td>fuji (Japanese wisteria)</td>
<td>7</td>
<td>54</td>
</tr>
<tr>
<td>khaki</td>
<td>7</td>
<td>22</td>
</tr>
<tr>
<td>momo (peach blossom)</td>
<td>6</td>
<td>113</td>
</tr>
<tr>
<td>olive</td>
<td>5</td>
<td>15</td>
</tr>
<tr>
<td>emerald</td>
<td>4</td>
<td>63</td>
</tr>
<tr>
<td>beni (cosmetic red)</td>
<td>4</td>
<td>30</td>
</tr>
<tr>
<td>sakura (cherry blossom)</td>
<td>4</td>
<td>15</td>
</tr>
<tr>
<td>uguisu (Japanese nightingale)</td>
<td>4</td>
<td>10</td>
</tr>
</tbody>
</table>

Table 1. Color terms, their common meanings, and usage across subjects and samples.

Figure 1. Histogram of color-term usage by our subjects. The list includes all terms used by four or more subjects.
segment for nonbasic terms. Particularly, *hada* fell on this declining segment.

**Japanese color glosses**

After excluding achromatic colors, 828 vectors derived from all 57 subjects’ color-term responses were used for *k*-means and gap-statistic analyses. Figure 4 shows the results of gap-statistic calculations for *k* = 2, . . . , 25, with 1,000 Monte Carlo simulations for each value of *k*. There is some variation in the gap-statistic analyses, because variations in the sums of squared intracluster distances were associated with slightly varying *k*-means solutions for the data and the reference data sets. The consistency and stability of *k*-means solutions are discussed in Appendix A. The inset histogram shows the frequency of *k*ₜₐₓ, as determined by the gap statistic (see L&B, equations 1–3); choosing the modal value of *k*ₜₐₓ for this distribution yielded *k*ₜₐₓ = 16.

Consensus maps for the 16 chromatic categories that were derived from our cluster analysis are shown in Figure 5, which are identified by their Japanese terms in the present study and their corresponding English color terms from L&B where possible. These maps were generated by element-wise summation of the binary vectors across subjects for each category, followed by scalar division by the maximum element for that category. The first eight of these consensus maps (left two columns) correspond to the eight universal chromatic color categories of Berlin and Kay (Kay et al., 2009; Lindsey & Brown, 2006): *aka* (*red*), *ki* (*yellow*), *midori* (*green*), *ao* (*blue*), *pink*, *orange*, *cha* (*brown*), and *murasaki* (*purple*), all of which (except for *ao*) are similar in location and extent to their American English counterparts (see Discussion for further details). The right two columns show eight additional color categories—*mizu* (*light blue*), *hada* (*skin tone*), *oudo* (*light brown*), *kon* (*dark blue*), *cream*, *matcha* (*yellowish green*), *enji* (*maroon/burgundy*), and *yamabuki* (*gold*)—many of which also have similar representations in American English. By adding three achromatic color terms, which gloss to *black*, *white*, *gray* (Figure 5A), *red*, *yellow*, *green*, *blue*, *pink*, *orange*, *brown*, *purple*, *light blue*, *skin tone*, *light brown*, *dark blue*, *cream*, *yellowish green*, *maroon/burgundy*, and *gold*, the full 16-chromatic category map was created (Figure 5B).

---

*Figure 2. The distribution of the number of color terms used by each of 57 Japanese subjects. Note: One subject used just the 11 basic color terms of Berlin and Kay (1969).*

*Figure 3. Color-term popularity diagrams. Data points are false-color coded by their meanings. Triangles: the basic color terms of Berlin and Kay (1969); circles: nonbasic terms; square: *mizu*, likely a basic color term in modern Japanese. (A) Sorted popularity of the unanalyzed color terms. (B) Sorted popularities of the color categories revealed by *k*-means analysis. Diagrams were fitted descriptively by a bilinear equation using a least-squares criterion. The descending limbs had slopes of −3.12 (A) and −4.114 (B).*

*Figure 4. Gap-statistic analysis of the *k*-means results on the raw color-naming data, with results of 1,000 tests. Following Tibshirani et al. (2001), the minimum value of *k* for which the gap statistic *G*(*k*) ≥ 0 and *G*(*k* + 1) < 0 (see Lindsey & Brown, 2014, equation 3). The inset shows that 16 is the modal value of *k*ₜₐₓ for this distribution of test results. See text for further details.*
and gray, the optimal number of categorical color terms in Japanese was 19.

Figure 3B shows a log-log plot of the popularity of the glossed color terms, sorted by rank. This plot is quite similar to the corresponding plot of raw color-term usage in Figure 3A, with one curiosity: In the glossed-usage rank plot, the ao category is used by only 56 subjects, whereas the mizu category is used by all 57 informants; this is reversed in the raw rank-usage plot, where all 57 subjects used ao while only 55 used mizu. This occurred because one informant used ao to name all the bluish samples, and the k-means algorithm assigned this ao feature vector to the mizu cluster. Additionally, one subject used the mizu synonym sora (sky) exclusively; k-means glossed this to the mizu cluster.

The Japanese color motifs

Motif analysis revealed two statistically significant motifs, shown in Figure 6. Median percentages of samples associated with each glossed term and the centroids for each corresponding category (in WCS coordinates) are tabulated in Appendix B. Inspection of the consensus maps for these two motifs and the distribution of the glossed color terms used by a plurality of subjects reveals little difference between them, except in the extensions of the mizu (light blue) and hada (skin tone) categories across the WCS color chart (arrows). Kay et al. (2009) suggest a threshold of 80% or greater consensus for terms to be candidates for basicness in any particular language. That threshold is achieved for 12 of the 19 Japanese color categories (16 chromatic plus three achromatic) derived from our cluster analysis of the raw Japanese color-naming data. Ten of these conform closely to the corresponding English BCTs: kuro/black, shiro/white, hai/gray, aka/red, ki/yellow, midori/green, pink/pink, orange/orange, cha/brown, and murasaki/purple. The eleventh, ao/blue, deviates from American English in that the English blue covers all shades of light and dark blue, whereas ao generally means medium to dark blue. The twelfth Japanese color category, mizu (light blue), does not have an equivalent category in American English, and represents a fundamental difference between it and the Japanese language in the lexical representation of color.

There was no difference in age between the members of the first and second motifs—t(55) = 0.669, p = 0.506—which is not surprising considering the narrow overall range of informant ages. There was only a nonsignificant trend for the first motif to contain more male data sets and the second motif to contain more female data sets (p = 0.061, two tailed, on Fisher’s exact test).

The existence of two motifs among speakers of modern Japanese adds to the growing body of evidence...
for diversity in color idiolects within the language communities of modern industrialized societies. This diversity is not arbitrary: It is structured in such a way that a few motifs are used, each with a relatively high consensus, by different subpopulations of a language community. It is our view that this diversity is essential for the evolution of color lexicons: The number of BCTs in a language increases as more and more members of a language community adopt a motif that is a variant of the existing lexicon but that contains more color terms. Our cluster/gap-statistic analysis of modern Japanese provides a snapshot of this process at one point in time.

**Discussion**

This study investigated the characteristics, statistics, and color-term usage frequency of the Japanese color language. We used color-term usage frequency and k-means clustering with gap-statistic analysis to determine the number of basic and salient color categories in Japanese. We discovered two distinct motifs in Japanese color naming, which were distinguished primarily by usage of mizu. Finally, we compared Japanese and American English color categories using separate and combined k-means analyses. We add correlation analysis and group mutual information later.

**The evolution of Japanese color-term usage**

A primary goal of this project was to determine whether the Japanese color lexicon had changed significantly in the 30 years since U&B’s study. Although U&B and the present study used somewhat different palettes of color samples, certain trends are clear: Most color terms have changed little, but mizu has probably become a BCT, and a few nonbasic colors are named with new terms.

In general, the raw color-naming data in the two studies show similar results. The terms for most Japanese basic color categories have not changed: Aka, ki, midori, ao, pink/momo, orange/daidai, cha, and murasaki still correspond to red, yellow, green, blue, pink, orange, brown, and purple, and the achromatic color terms kuro, shiro, and hai still correspond to black, white, and gray. Also, the nonbasic terms hadn, oudo, yamabuki, and cream were all used similarly and with relatively high consensus, both here and in U&B. The use of the English loanwords pink, orange, and gray instead of the traditional Japanese terms momo, daidai, and hai, respectively, remains unchanged. Thus, at this level of analysis, the Japanese color lexicon has evidently changed little since U&B’s study.

However, our results diverge from those of U&B in the usage of the terms for blue (mizu, ao, and kon). U&B rejected mizu as a BCT, even though the maximum consensus for mizu was high (nine out of 10 subjects). They did so primarily because 77% of samples named mizu by some of their subjects were named ao (blue) by other subjects. Their tabulated data also show that 80% of the samples called sora (sky) were sometimes called ao. This suggested that mizu and sora were subsets of ao, and therefore not BCTs.

In contrast, we argue that light blue is now a basic color category among speakers of Japanese, and mizu is now a BCT. First, like U&B, we find that almost all subjects used mizu. Of the 57 subjects in the present study, 54 used mizu to name two or more color samples, one named a single sample mizu, and one used sora, a synonym for mizu, to name the light-blue samples. Thus, only a single subject out of 57 failed to use any distinct term for light blue. Second, mizu named a large, well-delineated, generally contiguous range of samples (about the same size as the pink category; see
samples called mizu or sora by at least one subject

Figure 7. The World Color Survey chart showing all samples named mizu or ao by at least one of the 57 Japanese-speaking subjects in the present study. False-colored regions indicate samples whose responses used color terms at consensus \( \geq 0.5 \). Red asterisks: samples named ao or mizu with highest consensus (0.98 and 0.89, respectively). Black dots: samples called sora (sky) by at least one subject.

Figure 5). The set of all samples that were ever called mizu contained a nucleus of color samples for which mizu itself was the most frequent color term (Figure 6A). It also contained many samples that were called by other color terms by a majority of subjects (Figure 7). Notice that, although sora was used infrequently in the present data set, its distribution matched that of mizu almost exactly (black dots in Figure 7).

Third, the present data set contains evidence that mizu is no longer a subset of ao (blue). To compare our results to those of U&B, we repeated their calculation of the fraction of samples called mizu or sora by some subjects that were called ao by others. To make our results comparable to the U&B results, we created 10,000 10-subject subsets of the 54 subjects who used mizu or sora to name two or more samples. In 68% of the 10-subject subsets, there was at least one color sample that was called mizu by all 10 subjects. The blue histogram in Figure 8 shows the distribution of the results of our simulation. The median value of the distribution was 0.59 (blue arrow and blue-ringed dot), which fell below the 95% confidence intervals around U&B’s values (mizu: [0.668, 0.851]; sora: [0.664, 0.885]; horizontal blue bars in Figure 8). Furthermore, U&B’s values for mizu (0.77) and sora (0.80; solid blue dots in Figure 8) fell in the 96th and 98th percentiles of our distribution, respectively. Thus, our results are statistically significantly below those of U&B in Figure 8.

All these results agree with our motif analysis in suggesting that the Japanese language now includes a light blue color category. The results were identical (within rounding error) when we included only the mizu data (line histogram in Figure 8), suggesting that mizu is now a BCT in its own right. We propose that Japanese be added to the list of world languages that include a distinct BCT for light blue. In contrast to English, light blue has been reported to be a standard (perhaps basic) color term in some Indo-European languages—Russian (Paramei, 2005; Winawer et al., 2007), modern Greek (Androulaki et al., 2006; Thierry, Athanasopoulos, Wiggett, Dering, & Kuipers, 2009), Italian (Bimler & Uuskla, 2014), some forms of Spanish (Bolton, 1978; Harkness, 1973), Nepali (Bolton, Curtis, & Thomas, 1980), and Farsi (Friedl, 1979)—as well as some non-Indo-European languages: Turkish (Ozgen & Davies, 1998) and some forms of Arabic (Al-Rasheed, Al-Sharif, Thabit, Al-Mohimeed, & Davies, 2011; Borg, 2007).

Historical linguists (e.g., Stanlaw, 2010) often include the term kon (indigo) for dark blue among Japanese BCTs. Although kon is associated with one of the 16 chromatic categories derived by our cluster analysis, it was used by only 37% of subjects (Figure 1), which is fewer subjects than those who used either mizu or hada. U&B did not consider kon to be a possible BCT, because of all the color terms they listed, the consensus level for kon was the lowest (20%) and response time the second-longest (2.8 s). However, U&B’s centroid for kon in the OSA color sample space was not included in our Munsell stimulus set, because it was substantially less saturated than our corresponding samples (conversion charts from Nickerson, 1978).

Figure 8. Histograms of the outcomes of analyses of 10,000 random samples of 57 Japanese subjects containing exactly 10 subjects. Solid histograms: the fraction of outcomes in which a sample that was called x by at least one of the 10 subjects was also called y by at least one other subject. Light blue: \( x = \text{mizu}, y = \text{ao} \). Red: \( x = \text{aka}, y = \text{pink} \). Green: \( x = \text{midori}, y = \text{ao} \). Continuous-line histogram: \( x = (\text{mizu or sora}), y = \text{ao} \). Arrows and open circles: histogram medians. Solid circles and horizontal bars: fractions and 95% confidence intervals for the indicated color combinations from Uchikawa and Boynton (1987). Dotted lines join medians from the present study to the corresponding data from that study.

Also, our kon centroid was both darker and bluer (less purple) than that of U&B. In a related analysis, Stanlaw (2010) suggested that kon may be a necessary element of the modern Japanese color lexicon, and he reported that kon applies to blue samples that were very dark and somewhat desaturated. Thus, we tentatively agree with U&B that kon is not a BCT. However, we acknowledge that the best examples of kon may fall outside our test palette, so we cannot determine whether there has been a change in the status of kon over the last 30 years.

Turning to other regions of the color chart, U&B compare their results for ao versus mizu to their data for momo/pink versus aka (red). Because only 60% of the samples called aka were also called momo or pink (solid red dot in Figure 8), U&B concluded that momo/pink, unlike mizu, was a BCT. We performed a parallel analysis on the present data set, examining the data from those subjects who named two or more samples aka to determine what fraction of the aka samples were also called momo or pink. The median result from our 10,000 10-subject subsamples was 0.45 (red arrow and red-ringed dot), which fell below the 95% confidence interval around U&B’s 0.60 [0.462, 0.724] (horizontal red bar in Figure 8), and U&B’s value of 0.60 fell in the 99th percentile of our aka/pink distribution (red histogram in Figure 8). Thus, the present data set suggests that the momo/pink color category now covers an even more distinct region in color space than it did in the U&B data set. Therefore, our results concur with theirs that momo/pink is indeed a distinct color category and therefore a BCT. For comparison to mizu/ao and aka/pink, Figure 8 (green histogram) also shows that midori is not a subset of ao in the present data set.

There are also some differences between U&B’s data and the present data set in the words themselves: Few subjects in the present study used kusa (grass) to refer to yellow-green samples, preferring instead the term matcha (ceremonial green tea). Also, azuki (red beans) is now more commonly called enji (maroon).

In a summary of his work with the OSA color set, Boynton (1997) speculated that there might be room for a new universal color category in the area variously called peach or tan in English. Like U&B, we find common use of hada (skin tone) to denote colors in this range, but our cluster analysis also found that some subjects used it to mean cream. In showing a bimodal distribution of deployment, hada is similar to tan in American English, which some informants used to label pale-orange samples while others used it to label olive samples.

Hada is often used to name the foundations of Japanese cosmetics, and is probably more specific than flesh is in English, because the range of variation in the skin tone has been smaller in Japan for a long time. The frequency of hada usage observed in the present data set (48/57) is interesting because it indicates that hada is still in common use in everyday life, even after an educational campaign to discourage its use on social occasions because of possible racial connotations. In the present data set, hada was used nearly as frequently as the Japanese BCTs and mizu, and our rank-order-statistic analysis (Figure 2) suggests that its popularity is more in line with the BCTs than with the other nonbasic terms we encountered in our study. Future study of hada will reveal whether it continues to occur frequently.

Japanese is not a grue language

The oldest surviving written records in the Japanese language of ordinary people are the Manyō-Shū poems, which date from approximately 750 (original text: Frellesvig, Wright, Russell, & Sells, 2016; available in translation: McCauley, 2001a; reviewed in English: Stanlaw, 2007, 2010; Conlan, 2005). The color terms ao (blue) and midori (green) existed side by side in this early period, and their uses were not well distinguished. For example, in the Manyō-Shū poems, the color term ao (precursor to the modern ao) was sometimes used to name the colors of things that were clearly gray (a dappled gray horse [poem 136]), things that could be either green or blue (seaweed [poem 131]; mountains seen in the distance [poems 688, 923, 2707]), or things that were clearly green (leaves [poems 16, 2177]; grass [poem 2540]). Ao was also used frequently in the phrase “(青丹青)” (“auspicious blue, vermilion [red] clay”) as a metaphorical, honorific reference to the capital city Nara (over a dozen poems, including poems 29, 128, 1046, which are available in translation [McCauley, 2001a], and poem 328, which is most famous [Haitani, 2007]). Ao was also used for clouds (e.g., poem 3329; Nippon Gakujutsu Shinkokai, 1965, pg. 311) which were probably intended to be blue rather than green. Midori was used less frequently in Old Japanese, but it too sometimes meant blue, sometimes green. In a poem dating from 1192, midori named the sky (midori-no-sora), which was clearly blue (McCauley, 2001b), but in Manyō-Shū poems 2177 and 2540, midori named the color of summer leaves and grass, respectively, which were clearly green. Whereas ao continued to be used to name some green things as well as blue things, and sometimes black and gray things as well, over the next centuries midori became restricted to green. During the Meiji era (1868–1912), ao generally came to mean blue rather than green (e.g., a poem by Wakayama Bokusui, translated by Rimer & Gessel, 2005, pg. 311), but in a holdover from its previous meanings, ao is still used today as the color of fresh green shoots and the green traffic light. Contrary to these specific instances where ao denotes certain
green things, the present results and the results of other studies indicate that modern Japanese is not a grue language: Like modern English, the usual word for blue (ao) covers only blue samples (Figures 5 and 6, and the green histogram in Figure 8), and phrases like midori-no-sora sound strange to the modern ear. The modern Japanese use of ao to name certain specific green things is best attributed to custom and cultural connotation, perhaps analogous to the English use of blue to name the color of the blood of an aristocrat.

In an interesting parallel to the development of Japanese, Old English had both a term for grue (hœwen) and a term for green (grene; reviewed by Biggam, 1997). Hœwen meant green, blue, gray, and possibly purple (Clark Hall, 1916, pg. 144) among the common people, but it was restricted to blue among the educated elite. Bleu moved from Norman French into Middle English in the 13th century, replacing hœwen and meaning only blue, whereas the meaning of grene changed little. In modern English, blue and green are distinct color terms. It is important to recognize, however, that Japanese and English lexical divisions of blues and greens have been changing more recently in different ways. For example, the basic Japanese color term mizu (light blue) has no common equivalent in American English (L&B), whereas Japanese lacks a term that is equivalent to teal, a common American English color category that straddles the boundary between blue and green.

Comparison of Japanese and English color lexicons

It has long been known that the Japanese and English color-naming systems have similar structures (Berlin & Kay, 1969). Our cluster analysis confirms this basic result, although quantitative analysis of the clusters does reveal 16 chromatic clusters in Japanese, as compared to 17 in American English (L&B). Seven of the 16 Japanese clusters are similar to color categories in American English (L&B), as well as the universal color categories derived from the WCS by Lindsey and Brown (2006, 2009). To quantify these similarities, we computed correlation coefficients for the seven pairs of consensus maps for the Japanese/English clusters: aka/red, ki/yellow, midori/green, pink/pink, orange/orange, cha/brown, and murasaki/purple. All those correlation coefficients were >0.95. The correlation coefficient for the eighth pair, ao/blue, was only 0.79. This is not surprising given that lighter and less saturated colors in Japanese fall into the mizu category, which has no common English equivalent. Other highly correlated Japanese/English category pairs are enji/maroon (r = 0.91) and hada/peach (r = 0.76). Consensus maps for other Japanese and American English color terms have a good deal of overlap but generally do not match as well as the pairs listed already. For example the correlation coefficients for matcha/olive and yamabuki/gold were both 0.65, which was only slightly higher than those for oudo/olive (0.52) and oudo/gold (0.57). Moreover, L&B did not find an American English term corresponding to kon (dark blue). Conversely, our analysis did not reveal any Japanese terms corresponding to American English teal, magenta, lavender, or lime, which were derived by cluster analysis in L&B.

The present analysis indicates that there are two statistically significant color-naming systems, or motifs, in the Japanese language; and L&B revealed two motifs in American English. However, the motifs found in the two languages are clearly different. The Japanese motifs differ from each other mainly in the extension of mizu, while the American English motifs differ qualitatively, in that one contains high-consensus terms corresponding to the 11 universal color categories of Berlin and Kay while the other includes the high-consensus, nonbasic terms maroon and peach (equivalents of which are found in Japanese) as well as teal and lavender (which have no Japanese equivalents).

We compared Japanese and English color lexicons in another way, by pooling the two data sets and clustering them simultaneously to reveal a single set of glossed color categories and a single set of motifs. Our first k-means cluster step revealed 11 glossed chromatic color categories that were similar to corresponding categories in both Japanese (Figure 5) and American English (L&B, figure 9). These are the eight chromatic color terms of Berlin and Kay, plus hada/peach, matcha/olive, and enji/maroon (Figure 9, panels with both Japanese and English terms). It also revealed six glossed color categories that are well represented in one language but not the other (panels with single terms). Our second analysis revealed three motifs, all of which contained the eight universal chromatic color categories of Berlin and Kay plus black, white, and gray. The first of the three pooled motifs (Figure 10A) contained mainly English data sets and featured a robust representation of Berlin and Kay’s universal color categories only. The second (Figure 10B) contained mainly Japanese data sets, and added high-consensus mizu to the universal color categories. The third motif, which contained mainly English data sets (Figure 10C), added the high-consensus English categories teal, lavender, and peach.

Color communication in Japanese and American English

In the present study, Japanese subjects used fewer terms than the American subjects did in L&B, and...
there was also one fewer lexical category in Japanese ($k_{L,opt,Jap} = 16$; Figure 5) than in American English ($k_{L,opt,Eng} = 17$; L&B, figure 3), as revealed by cluster analyses. Moreover, although these analyses revealed two motifs for both the Japanese and English lexicons, the English motifs were more distinct from each other than the Japanese motifs were, with the two Japanese motifs differing only in how far *mizu* extends across the color chart. These comparisons suggested that Japanese color naming might be more consistent across subjects than English color naming is.

We therefore compared color-naming consistency in Japanese and American English, using an information-theoretic metric for color communication—group mutual information (GMI)—described by Lindsey et al. (2015). GMI quantifies information transfer (in bits) within a language community, based on each informant’s color names for a standard set of colors. It differs from other metrics for consistency, such as

---

**Figure 9.** Consensus maps for the 17 chromatic color terms revealed by the $k$-means and gap-statistic analysis of the pooled Japanese and American English color-naming data. See text for further details.

**Figure 10.** Three motifs based on pooled data from 57 Japanese-speaking and 51 American English–speaking informants. Conventions as in Figure 6. At the 80% consensus criterion, Motif 1 ($N = 40$: four Japanese and 36 English): *aka/ red, ki/yellow, midori/green, ao/blue, pink/pink, orange/orange, cha/brown, murasaki/purple, shiro/white*; Motif 2 ($N = 53$: 51 Japanese and two English): *mizu* added; Motif 3 ($N = 15$: two Japanese and 13 American English): with *teal* and *lavender* but no *mizu* or *shiro/white*.
consensus, because it takes into account both the agreement across subjects in the deployment of a given lexicon and the number of different lexical color categories, and the relative extents of these categories across the chosen color palette.

The GMIs for Japanese and American English, calculated from the raw color-naming data for each language, were 2.04 and 1.83 bits, respectively. These GMIs are substantially lower than the 3.46 bits expected from Berlin and Kay’s 11 BCTs when the corresponding color terms are deployed optimally and consensus across speakers is perfect (Lindsey et al., 2015). If we were to assume that all 32 terms deployed by four or more Japanese subjects were deployed optimally and with perfect consensus, then the GMI would be 5.0 bits. The differences between optimal and measured GMI reflect diversity across individuals in the number of named color categories, the choice of color terms, and the locations and sizes of these categories within color space. We also calculated the GMIs for Japanese and American English based on the glossed color-category clusters at their respective values of \( k_{L, opt} \). When the synonyms in both languages are consolidated by cluster analysis, the GMI for Japanese (2.28) is still higher than the GMI for American English (2.15). Thus, it appears that that Japanese color terms are deployed more consistently across individuals, producing better GMI, even though Japanese speakers have one fewer salient color term than English speakers do. This trend is also reflected in the steepness of the line fitted to the second segment in the log-log plot of rank-order analysis (Figure 3A). Based upon these results, we hypothesize that Japanese speakers are more efficient at color communication than speakers of American English.

**Summary**

We studied color naming in 57 native Japanese-speaking subjects using a diverse palette of 330 Munsell color samples that was used previously in the World Color Survey and in a recent study of American English (Lindsey & Brown, 2014). Japanese subjects used a total of 93 terms to name the color samples. Thirty-two of these terms were used by four or more subjects. Our \( k \)-means/gap-statistic analyses suggest that there are no more than about 19 distinct lexical color categories in the Japanese language.

Comparison of the present results with those of a prior study by Uchikawa and Boynton (1987) revealed substantial similarities between the two studies. However, there were also differences, which suggested that the Japanese color lexicon has changed somewhat in the last 30 years. Notably, *mizu* (*light blue*) is likely emerging as a new basic color term. *Kon* (*dark blue*) appears more often and *kusa* (*yellow-green*) less often here than in that study. It is important to keep in mind that the two studies used different color palettes, and this may account for some of the differences in results between them.

Comparison between the Japanese and American English color lexicons revealed significant similarities as well as differences between the two languages. Our \( k \)-means analysis showed that consensus maps for the eight chromatic basic color terms of Berlin and Kay (*red*, *yellow*, *green*, *blue*, *pink*, *orange*, *brown*, and *purple*) were strikingly similar across the two languages, with the exception of *blue*, which is truncated in extent in Japanese by the presence of *mizu* (*light blue*) and *kon* (*dark blue*). Consensus maps for the minority color categories were less well correlated across the two languages. Moreover, some categories present in one language (*mizu*, *kon*, *lavender*, *teal*, *magenta*, and *lime*) were not present in the other language.

Cluster analysis revealed two motifs both in Japanese and in American English. The two Japanese motifs differed only in how many *light blue* samples were called *mizu*, whereas the American English motifs reported by Lindsey and Brown (2014) were distinguished by the use of *teal*, *lavender*, *peach*, and *magenta* in one motif but not the other. Here we found no evidence in either the raw color-naming data or the results of our cluster analysis for the premodern usage of *ao* as a grue (*green-or-blue*) color term.

Information-theoretic analyses based on group mutual information reveal greater consistency in color naming across Japanese subjects as compared to American English subjects, regardless of whether the raw terms or \( k_{L, opt} \) glossed terms were used in the analysis.

**Keywords:** color category, color name, \( k \)-means analysis, motifs, Japanese color terms

**Acknowledgments**

This work was supported by a Cooperative Research Project (International type; No. H26/A25) of the Research Institute of Electrical Communication, Tohoku University to KU, JSPS KAKENHI No. JP15H03460 to IK, and NSF BCS-1152841 to DTL. We thank Moe Seto for assistance in collecting data at Tohoku University.

Commercial relationships: none.
Corresponding author: Ichiro Kuriki.
Email: ikuriki@riec.tohoku.ac.jp.
Address: Research Institute of Electrical Communication, Tohoku University, Sendai, Japan.
Footnotes

1Here, gold is the English term for a golden color; it should not be confused with the surface appearance of the precious metal.

2Let $a_{ij}$ be the number of times sample $i$ is assigned to cluster $j$ in $N$ k-means runs. Assume $i$ is assigned to a total of $q_i$ different clusters in $N$ runs. It follows that $q_i/N$ are frequencies of assignment and $\Sigma a_{ij}/N = 1.0$. Thus, if we calculate CI$(i)$ as the average frequency of cluster assignment across $q_i$ clusters, $CI(i) = (\Sigma a_{ij}/N)/q_i = 1/q_i$.

References


Appendix A: Stability of k-means solutions

It is known that k-means clustering can yield somewhat different solutions across runs, since k-means is an iterative algorithm designed to minimize the sums of total within-cluster differences across k clusters. Each k-means iteration begins with a set of k randomly assigned centroids, and different initial centroids can lead to solutions that are influenced by minimization of within-cluster differences around different local minima, especially when data sets are not particularly large.

The k-means analysis described in the main text revealed a specific set of 16 chromatic categories. It is natural to ask whether we would have obtained a substantially different set of chromatic categories on a different repetition of our k-means analysis of the Japanese color lexicon. In this appendix, we describe an examination of the consistency and stability of k-means clusterings of our color-naming data across multiple k-means runs.

We defined a consistency index CIk, which specified the consistency with which, for a given value of k, the k-means algorithm assigned each of the 320 chromatic samples to the same cluster across n = 100 runs of the algorithm. Each run consisted of 100 repetitions of k-means clustering, and the solution for that run was the repetition that produced the tightest cluster result, as determined by the sums of squared intracluster distances.

To evaluate the consistency of solutions across runs, each cluster in a given run was first uniquely “matched” to a single cluster in each of the other runs. Matching was accomplished by sorting the clusters in each solution by the number of color samples in each cluster, together with the correlation coefficients between cluster vectors across runs. In this way, a cluster identified with a particular gloss (say, for example, red) in a given solution was uniquely matched with one cluster in each of the other solutions that also corresponded to that gloss, or to the cluster with which it was most similar.

CIk was based on the frequencies with which the ith chromatic sample \( c(i) \) was assigned to the k matched
and nonmatched clusters, across the 100 $k$-means runs. Thus, if $c(i)$ were always assigned to, say, the red cluster, then $CI_k(i)$ would be 1.0. In general, if $c(i)$ were assigned to $q_i$ different clusters across the 100 $k$-means runs, then the average frequency in the assignment of $c(i)$ across the $q_i$ clusters is just:

$$CI_k(i) = 1/q_i. \quad (A1)$$

$CI_k$ is then the average of $CI_k(i)$ calculated for $i = 1, \ldots, 320$ chromatic samples:

$$CI_k = \frac{CI_k(i)}{i=320}. \quad (A2)$$

It is important to recognize that inconsistency does not generally occur at the level of the individual color sample. Instead, it appears when groups of color samples associated with individual color terms are assigned to different categories from one run to the next. This variation is due to variation in the random assignments of centroids at the start of each invocation of the $k$-means clustering algorithm. Figure A1B shows the results for the $k$-means solutions at $k = 5$, which gave the lowest CI in our analysis (see later). The inconsistencies at $k = 5$ appear in the form of large contiguous regions of the color chart where color samples have been assigned with somewhat low consistency; otherwise, consistency is uniformly at or near 1.0.

The process of computing $CI_k$ for 100 $k$-means runs was repeated 1,000 times, and the means and 95% confidence intervals were obtained for each value of $k$.

The results of average consistency for $k = 2, \ldots, 16$ are plotted in Figure A1B. The horizontal axis shows the number of clusters $k$, and the vertical axis shows the consistency index, averaged across repetitions. The consistency of clusters across runs was high for $k$ from 8 (which gives us 11 color categories) through 16 (the optimum number of categories revealed by our gap-statistic analysis). The consistency drops at $k = 4, 5, 6,$ and 7, with a minimum at $k = 5$, but the lower 95% confidence interval around the mean consistency index was above 0.9 otherwise, even out to and including $k = 16$, which we determined to be optimal based on our gap-statistic analysis.
### Appendix B. Median percentage of samples named by category (interquartile range) and median centroid (in World Color Survey notation).

**Notes:** *Median category usage differs significantly between motifs. †Median category centroid hue differs significantly between motifs. ‡Median centroid value differs significantly between motifs.*

<table>
<thead>
<tr>
<th>Ki</th>
<th>Matcha</th>
<th>Midori</th>
<th>Cha</th>
<th>Pink</th>
<th>Yambuki</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motif 1</td>
<td>4.4% (3.3%–5.5%), F06</td>
<td>0.0% (0.0%–1.4%), D14</td>
<td>25.0% (22.4%–27.4%), E19</td>
<td>6.7% (6.1%–7.6%), C08</td>
<td>10.2% (8.5%–11.2%), G01</td>
</tr>
<tr>
<td>Motif 2</td>
<td>4.6% (4.2%–5.6%), H11</td>
<td>0.0% (0.0%–0.3%), E14</td>
<td>23.6% (21.1%–25.8%), E19</td>
<td>6.7% (5.7%–7.9%), C08</td>
<td>8.2% (6.4%–9.2%), G01</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Orange</th>
<th>Ao*†‡</th>
<th>Oudo</th>
<th>Enji</th>
<th>Mizu*</th>
<th>Aka</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.5% (3.8%–5.5%), F06</td>
<td>7.6% (6.7%–9.4%), D29</td>
<td>0.9% (0.0%–1.7%), E11</td>
<td>0.0% (0.0%–0.5%), C04</td>
<td>9.1% (7.9%–10.9%), H26</td>
<td>3.6% (2.6%–4.9%), D03</td>
</tr>
<tr>
<td>3.9% (3.2%–5.5%), F06</td>
<td>11.8% (10.3%–13.5%), E28</td>
<td>1.2% (0.0%–2.4%), E11</td>
<td>0.0% (0.0%–0.7%), C04</td>
<td>6.4% (4.2%–8.5%), H26</td>
<td>4.5% (3.3%–6.1%), D03</td>
</tr>
</tbody>
</table>

### Appendix B. Extended

<table>
<thead>
<tr>
<th>Murasaki</th>
<th>Kon</th>
<th>Hada (light)</th>
<th>Hada (dark)</th>
<th>Kuro</th>
<th>Hai</th>
<th>Shiro*</th>
</tr>
</thead>
<tbody>
<tr>
<td>13.0% (11.8%–14.2%), D35</td>
<td>0.2% (0.0%–1.5%), B29</td>
<td>0.0% (0.0%–0.9%), I11</td>
<td>2.3% (1.4%–3.6%), H06</td>
<td>2.1% (1.1%–2.9%), H06</td>
<td>1.7% (1.2%–2.3%), 3.6% (3.0%–5.5%)</td>
<td></td>
</tr>
<tr>
<td>13.3% (10.9%–14.2%), D36</td>
<td>0.0% (0.0%–0.0%), B29</td>
<td>0.0% (0.0%–0.4%), I12</td>
<td>2.1% (1.4%–3.1%), H06</td>
<td>2.7% (2.1%–3.0%), H06</td>
<td>1.5% (1.2%–1.6%), 6.7% (5.2%–9.2%)</td>
<td></td>
</tr>
</tbody>
</table>

Appendix B. Extended