

# Basic color categories in Mandarin Chinese revealed by cluster analysis

**Tsuei-Ju Tracy Hsieh**

Department of Information Communication,  
Chinese Culture University, Taipei, Taiwan



**Ichiro Kuriki**

Research Institute of Electrical Communication,  
Tohoku University, Sendai, Japan



**I-Ping Chen**

Institute of Applied Arts, National Chiao Tung University,  
Hsinchu, Taiwan



**Yumiko Muto**

Tokyo Institute of Technology Graduate School,  
Yokohama, Japan  
Brain Science Institute, Tamagawa University,  
Tokyo, Japan



**Rumi Tokunaga**

Graduate School of Global and Transdisciplinary Studies,  
Chiba University, Chiba, Japan



**Satoshi Shioiri**

Research Institute of Electrical Communication,  
Tohoku University, Sendai, Japan



Previous claims of the number of color categories and corresponding basic color terms in modern Mandarin Chinese remain irreconcilable, mainly due to the shortage in objectively evaluating the basicness of color terms with statistical significance. Therefore the present study applied *k*-means cluster analysis to investigate native Mandarin Chinese speakers' color naming data of 330 color chips similar to those used in World Color Survey. Results confirmed that there are 11 basic color categories among modern Mandarin speakers in Taiwan, one corresponding to each basic color term. Results also showed that observers overwhelmingly agreed in their use of Mandarin color terms, including those that had yielded ambiguous results in previous studies (gray, brown, pink, and orange). There is significant cross-language similarity when comparing the distribution of color categories in the World Color Survey chart with American English and Japanese data. The motif analysis and group mutual information analysis suggest that Mandarin color terms used in Taiwan describe very similar categories and are, hence, similarly precise in communicating color information as those in Japanese and American English. These results show that three languages of fundamentally different cultures and histories have very similar basic color terms.

## Introduction

Research on the phenomenon of language-dependent color perception began decades ago (e.g., Whorf, 1956; Brown & Lenneberg, 1954). Theories and methodologies around color naming and color categorization are still evolving (e.g., Lindsey & Brown, 2019; Mylonas & Griffin, 2020; Siuda-Krzywicka, Witzel, Bartolomeo, & Cohen, 2020). One of the essential issues that received much attention is cross-cultural regularity in the categories that correspond to basic color terms, namely basic color categories (e.g., Berlin & Kay, 1969; Lindsey & Brown, 2009; Kay, Witzel, Bartolomeo, & Cohen, 2011), because the empirical findings are influential to the classical debate between universalistic and relativistic relationships between color term and color category (for overview, see Jraissati, 2014; Heider, 1972; Lindsey & Brown, 2019; Witzel, 2019). Various languages have been examined with respect to the cross-culture regularity of basic color categories through commonly used procedures such as the World Color Survey (for an overview see Kay, 2015). Mandarin is an interesting case for investigating differences in color naming across languages because its history and culture differ from

Citation: Hsieh, T.-J. T., Kuriki, I., Chen, I.-P., Muto, Y., Tokunaga, R., & Shioiri, S. (2020). Basic color categories in Mandarin Chinese revealed by cluster analysis. *Journal of Vision*, 20(12):6, 1–17, <https://doi.org/10.1167/jov.20.12.6>.



BCT studies	N	Method	English translation
<b>Berlin &amp; Kay (1969)</b>	6	紅 hóng, 黃 huáng, 綠 lǜ, 藍 lán, 黑 hēi, 白 bái red, yellow, green, blue, black, white	Color naming task with Chinese immigrants in US
<b>Lu (1997)</b>	11	紅 hóng, 橙 chéng, 黃 huáng, 綠 lǜ, 藍 lán, 桃 táo, 紫 zǐ, 褐 hé, 灰 huī, 黑 hēi, 白 bái red, orange, yellow, green, blue, pink, purple, brown, gray, black, white	Free-naming method and color naming task with Taiwanese participants
<b>Lin et al. (2001a; 2001b)</b>	11	紅 hóng, 橘 jú, 黃 huáng, 綠 lǜ, 藍 lán, 粉紅 fěnhóng, 紫 zǐ, 棕 zōng, 灰 huī, 黑 hēi, 白 bái red, orange, yellow, green, blue, pink, purple, brown, gray, black, white	Free color naming task with 40 Taiwanese participants
<b>Hsieh &amp; Chen (2011)</b>	11	紅 hóng, 橘 jú, 黃 huáng, 綠 lǜ, 藍 lán, 粉紅 fěnhóng, 紫 zǐ, 咖啡 kāfēi, 灰 huī, 黑 hēi, 白 bái red, orange, yellow, green, blue, pink, purple, brown, gray, black, white	Free-naming method and color naming task with Taiwanese participants
<b>Wu (2011)</b>	8	紅 hóng, 黃 huáng, 綠 lǜ, 藍 lán, 紫 zǐ, 灰 huī, 黑 hēi, 白 bái red, yellow, green, blue, purple, gray, black, white	Word frequency accumulation of modern era from a Chinese Corpus
<b>Gao &amp; Sutrop (2014)</b>	9	紅 hóng, 黃 huáng, 綠 lǜ, 藍 lán, 粉紅 fěnhóng, 紫 zǐ, 灰 huī, 黑 hēi, 白 bái red, yellow, green, blue, pink, purple, gray, black, white	Free-naming method and color naming task with Northeast China participants
<b>Sun &amp; Chen (2018)</b>	8	紅 hóng, 棕 zōng or/and 褐 hé, 黑 hēi, 橙 chéng, 灰 huī, 黃 huáng red, brown, green, black, orange, gray, yellow, pink	A matching task of 32 terms to color chips with Taiwanese participants

Table 1. Previous studies about the numbers of BCTs in modern Mandarin Chinese.

languages whose basic color terms are well known, such as English and Japanese (Lindsey & Brown, 2014; Kuriki, Lange, Muto, Brown, Fukuda, Tokunaga, Lindsey, Uchikawa & Shioiri, 2017). Many attempts have been made to determine the basic color terms of Mandarin.

Table 1 summarizes several studies on basic color terms (BCTs) of modern Mandarin by listing each claimed number of BCTs (*N*), the BCTs in Chinese characters and corresponding translation in English, and each applied method. The study by Berlin & Kay (1969) recognized only six BCTs referring to red, yellow, green, blue, black, and white. A few later studies stated the imperfections in Berlin and Kay's survey and presumed that there should be 11 BCTs in Mandarin Chinese, but their surveys resulted in different BCTs within specific categories such as brown (Lu, 1997; Lin, 2001; Hsieh & Chen, 2011). Through the approach of the diachronic linguistic corpus, Wu (2011) quantified eight modern Mandarin BCTs, whereas Gao and Sutrop (2014) identified nine based on the criteria of BCTs proposed earlier (Berlin & Kay,

1969). These two studies claimed numbers of BCTs fewer than 11, and both excluded brown and orange as basic color categories. Wu (2011) excluded pink due to the absence of a monolexeme pink term in modern Mandarin Chinese. These three excluded categories presented multiple inconsistent terms in studies that claimed 11 Mandarin BCTs. Specifically, both “橙 chéng” (Lu, 1997) and “橘 jú” (Lin, Luo, MacDonald, & Tarrant, 2001a; Hsieh & Chen, 2011) were found to be orange BCTs. Both “桃 táo” (Lu, 1997) and “粉紅 fěnhóng” (Lin et al., 2001; Hsieh & Chen, 2011; Gao & Sutrop, 2014) were pink BCTs. Brown was especially controversial as it could be “褐 hé” (Lu, 1997), “棕 zōng” (Lin, Luo, MacDonald, & Tarrant, 2001b), or a foreign loanword meaning coffee “咖啡 kāfēi” (Hsieh & Chen, 2011).

Sun & Chen (2018) ascribed the discrepancy to the language characteristic of having multiple synonyms in Mandarin color terms, therefore, they adopted a rather novel method that combined the corpus data and lab-based color-matching task. They selected 32 traditional monolexeme color terms from a Chinese

linguistic corpus and asked their participants to pick color chips from the stimuli set similar to the World Color Survey that match each term. They used a principal component analysis to extract eight factors to conceptually represent a “family” of color terms for red, brown, green, black, orange, gray, yellow, and pink (see figures 4–12 in Sun & Chen, 2018 for details). However, the commonly identified categories (white, blue, and purple) did not emerge as influential factors from their analysis. Besides, using color terms from historical literature rather than direct naming could detract the comparability with other modern studies on Mandarin BCTs that applied the naming task approach.

To summarize, two open questions still remain: (1) how many and which basic color terms are there in Mandarin, and (2) how are they similar to languages such as English and Japanese? Let us begin with assessing why the previous studies yielded such disagreements. Considering that the determinants of BCTs are complicated (e.g., Hardin, Hardin, & Maffi, 1997; Biggam, 2012), one might suspect that the varied data sources or collection procedures could affect each result. Nevertheless, we are convinced that the main cause of the conflict claims is the shortage of rational and objective methods to evaluate the basicness of the Mandarin BCTs, and consequently the claimed numbers of the BCTs varied across studies.

Pertinent studies commonly quoted four compulsory criteria proposed by Berlin & Kay (1969), together with a few supplementary criteria, to qualitatively determine the basicness of the claimed BCTs. The four criteria for recognizing a BCT include the following: (1) it is mono-lexemic, that is, its meaning is not predictable from the meaning of its parts; (2) its significance is not included in that of any other color term, that is, it is not a hyponym; (3) its applicability must not be restricted to a narrow class of objects; and (4) it must be salient for informants. The first three criteria address the linguistic features of a BCT, whereas the final criterion of saliency describes behavioral features in eliciting or in using a BCT, such as a tendency to occur at the beginning of elicited lists of color terms, stability in its reference, and occurrence in the idiolects of all informants (Berlin & Kay, 1969). With these classic criteria in addition to the other discussed features for qualifying BCTs (e.g. Kay; 1978; Saunders, 2000; Mylonas & Griffin, 2020), previous studies have spent a lot of effort defending or refuting which Mandarin BCTs were genuine, because in some cases the criteria would contradict each other and it was not clear which criterion prevailed. For example, Berlin and Kay (1969) stated that “Mandarin was a problematic case” partially because of the difficulty of determining a gray BCT (for detail, see pp. 41–42 in Berlin and Kay, 1969). They then precluded a possible gray BCT “灰 hu” because its occurrence violates their hypothesized chronological order of BCTs. With respects to these widely applied

criteria, some commonly reported BCTs (see Table 1) could be excluded, such as “橙 chéng,” “橘 jú,” and “棕 zōng,” because they are object names meaning orange, tangerine, and palm tree, respectively, and are not purely color names; “咖啡 kāfēi” because it is a foreign loanword meaning coffee, and “粉紅 fěnhóng” because it is a compound word rather than a monolexeme. Most determinations have tended to be based on the rationale of criteria involving language or linguistic features. However, because of the difference in language attributes between Mandarin and English, such an application could be inappropriate.

Overall, several basic color categories remain debatable within the issue of Mandarin BCTs due to the lack of objective means to quantitatively evaluate the basicness, consequently making the status of modern Mandarin an open question in the scope of Berlin & Kay (1969)’s theory of evolutionary stages. Although there were studies that tried to provide empirical data such as descriptive data of naming frequency (Lu, 1997; Hsieh & Chen, 2011), the results without further statistical analysis barely correspond with behavioral characteristics in the criterion of saliency. Therefore the present study aims to justify the number of color categories with each Mandarin BCT through an analysis method that particularly specializes in revealing the grouping pattern of color samples. The method we adopted was *k*-means clustering technique that was found promising to reveal the basic color categories independent of their labels (i.e., color terms) (Lindsey & Brown, 2006; Lindsey & Brown, 2009; Kuriki et al., 2017). The present study also conducted the “motifs” analysis for revealing the similarity of color naming systems among individuals across language groups, as well as revealing the dissimilarity among motifs (Lindsey & Brown, 2006, 2009; Kuriki et al., 2017). The use of these methods allowed us to directly compare Mandarin (Taiwan) data with Japanese and English (American) data, collected with the same procedure. Therefore the primary purpose of the present study is to apply both *k*-means clustering and motif analyses that have not previously been used in studying Mandarin BCTs to enable quantitative analyses and discussions of BCTs in modern Mandarin.

## Method

### Participants

Participants were 41 healthy college students of the Chinese Cultural University of Taipei who spoke Mandarin Chinese as their first language. They were all color normal and tested with Ishihara pseudo-isochromatic plates (Ishihara, 1918), and all had normal or corrected to normal visual acuity.

Their ages ranged from 19 to 26 years ( $M = 21.4$ ,  $SD = 1.84$ ). They were volunteers recruited from the College of Journalism and Mass Communication. The review committees of the Center for Research Ethics at National Taiwan Normal University approved the procedure of the experiment. All participants' consent was given in a written form.

## Materials and procedure

The color chips from the *Munsell Book of Color, Glossy Edition* ([www.munsell.com](http://www.munsell.com)) were mounted on cardboard covered with a medium gray sheet that corresponds to N5/ in Munsell notation. The set of 330 chips (320 chromatic + 10 achromatic chips) were identical pieces used in a study on the color categories in modern Japanese by some of the authors' group (Kuriki et al., 2017). The color chips were presented one-by-one in a standard light cabinet, VeriVide Color Assessment Cabinet CAC60 (VeriVide, Leicester, UK) and illuminated with VeriVide F20T12 / D65 Light Tube (CIE  $R_a = 97$ , CIE(x, y) = (0.3179, 0.3375), approximate illuminance on the tabletop was 301 lux, measured by a spectroradiometer CS-1000A/S/T (Konica Minolta, Tokyo, Japan); see Supplementary Figure S1 for the spectrum of the light source). The experiment was conducted in a laboratory at Chinese Culture University in Taipei, Taiwan. The participants entered the laboratory where the VeriVide standard viewing cabinet was placed in an otherwise dark environment. The lamp on the viewing cabinet was turned on half an hour before when the participant entered the laboratory for the light source stability. The participant sat in front of the viewing cabinet for at least 10 minutes to adapt to the light while listening to the experimenter's instructions and signing an agreement form of informed consent. During the experiment of the color naming task, the participants were shown 330 color chips that were presented in a pseudo-random order. The participants were shown each color chip for at least five seconds, then they named each color chip using a single term (monolexic restriction). The participants delivered responses orally and were recorded using a digital audio recorder, ICD-SX2000 (Sony, Tokyo, Japan) that was later transcribed by the experimenters for analysis. The experiment session was not interrupted unless the participants asked for short breaks, and during the breaks they were asked to stay in front of the VeriVide cabinet to maintain the adaptation state. The overall duration for a participant to complete the color naming task was about 60 minutes.

To specify the monolexic restriction, the participants were told not to use tone modifiers (e.g., pale, light, dark, etc.) or compound words (e.g., reddish-purple, yellowish-green, etc.). In practice, the participants would be suggested to use other

words instead only when the participants used a compound term, except for “粉紅 (*fěnhóng*; pink).” “粉紅 (*fěnhóng*)” has been a debatable case in the studies of Mandarin Chinese BCTs (e.g. Gao, 2014) because it is a most commonly used modern color term for naming colors in the pink category; meanwhile it is a compound word combining a tone modifier “粉 (*fěn*; faint or light)” with a known BCT “紅 (*hóng*; red).” To better justify the basicness of “粉紅 (*fěnhóng*; pink),” the separation of categories named with “粉紅 *fěnhóng*” and “紅 *hóng*” will be assessed with a quantitative measure, overlap ratio (Kuriki et al., 2017). This analysis was conducted for all pairs of basic color terms. The details of this measure will be described later in the analysis section.

## Analysis

Since we employed the same analytical techniques as the previous studies, the details of analysis are same as described in the original articles (Kuriki et al., 2017, p. 4; Lindsey & Brown, 2014, p. 13; Lindsey & Brown, 2006). The outline of the analysis is as follows. First, we generated binary vectors for the result of each color term for each participant. Each  $1 \times 330$ -dimensional vector represented a term used by a participant, and “1” was assigned to the vector element that corresponds to a color chip named with the term and ‘0’ otherwise. The number of color terms, that is, the number of vectors, were 478 in total from 41 participants.

In  $k$ -means clustering, the number of clusters has to be given a priori. We used gap statistics (Tibshirani, Walther, & Hastie, 2001) as the measure of optimization for the clustering analysis (Kuriki et al., 2017, p. 4; Lindsey & Brown, 2006, 2009; Lindsey & Brown, 2014, p. 13). With the same method used in the original studies (Lindsey & Brown, 2006, 2009), the reference data sets were prepared by randomly shifting the location of measured color categories for 320 chromatic color chips in the Munsell color space (Kay & Regier, 2003), with the left and right ends of the color space of the chromatic color chips assumed to be a continuous hue circle like a cylinder, together with a random flip in the lightness direction (for details, see Figure 2 in Kay & Regier, 2003; Kuriki et al., 2017, p. 4). After this procedure, the reference was generated for 320 chromatic color chips, achromatic chips and responses were excluded from analyses that used this method (analyses for Figures 2–6). The  $k$ -means method was applied to 353 vectors (number of all chromatic responses)  $\times$  320 elements (number of color chips).

Similar to our previous study (Kuriki et al., 2017), we also evaluated a consistency index for the classification results to justify the optimality of the  $k$  with another measure. Since  $k$ -means clustering is a

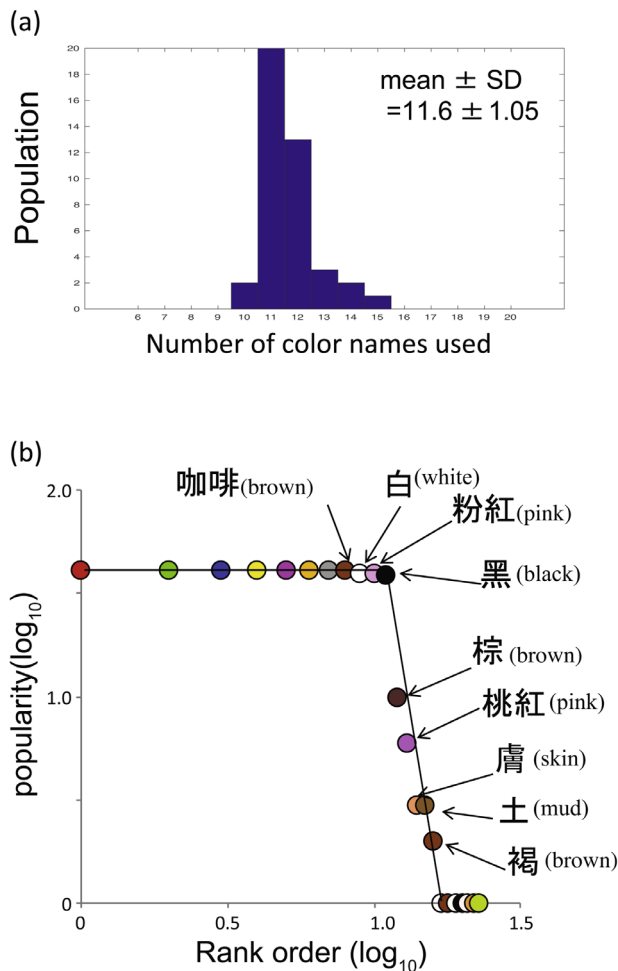


Figure 1. (A) Histogram of the numbers of color names used by Mandarin speakers in Taiwan. (B) Rank order plot (Zipf chart) of our Mandarin Chinese data. The *horizontal* and *vertical* axes represent logarithms of rank order and populations, respectively.

method that begins with randomly defined initial values (centroids), the final result could vary between runs. The reproducibility of each cluster was evaluated by calculating the correlation coefficient of vectors across repetitions, where similarity of clusters takes the value of 1.0 at the maximum.

In another type of  $k$ -means analysis to investigate the use of an achromatic term “灰 (*huī*; gray)” (Figure 7), we applied  $k$ -means analysis to 330-dimensional binary vectors (478 vectors, in total) using a different way of generating the reference data set. The difference will be described later when it arises.

To evaluate the distinctness between color categories obtained, especially between “紅 (*hóng*; red)” and “粉紅 (*fěnhóng*; pink),” we calculated the overlap ratio between color categories (Kuriki et al., 2017). This was defined by the relative number of color chips normalized to the total number of color chips named with a color term (e.g., “粉紅 (*fěnhóng*; pink)” in

Color terms

Rank	MC letter	Pronunciation	English	<i>N</i>
1	紅	hóng	Red	41
2	綠	lù	Green	41
3	藍	lán	Blue	41
4	黃	huán	Yellow	41
5	紫	zǐ	Purple	41
6	橘	jú	Orange	41
7	灰	huī	Gray	41
8	咖啡	kāfēi	Brown	41
9	白	bái	White	40
10	粉紅	fěnhóng	Pink	40
11	黑	hēi	Black	39
12	棕	zōng	Brown	10
13	桃紅	táohóng	Pink	6
14	土	tǔ	Mud	3
15	皮膚 膚	pífū, fū	Skin	3
16	褐	hé	Brown	2
17	澄	chéng	Clear	1
18	茶	chá	Brown	1
19	乳白	rǔbái	Milk	1
20	墨	mò	Ink	1
21	湛	zhàn	Pearly	1
22	橙	chéng	Orange	1
23	青蘋果	qīng píngguǒ	Green apple	1

Table 2. Color terms in Chinese characters, pinyin (pronunciation), English translations, and number of participants using them (*N*).

a participant that were also named by a different color term (e.g., “紅 (*hóng*; red)”) at least once by a different participant. To maintain the consistency of this measure to our previous study (Kuriki et al., 2017), we resampled the same number of participants, that is, 10, in the present study as well.

## Results

### Descriptive statistics

The number of color terms reported by each participant is shown in the histogram (Figure 1A). The number of color terms elicited from an individual participant ranged from 10 to 15: mean  $\pm$  SD =  $11.6 \pm .05$ . Figure 1B shows the popularity of each color name (Zipf analysis) and Table 2 the number of participants for each color term used.

The 11 basic color terms were all included in the most popular group ( $N > 38$ , 93%): “紅 (*hóng*; red),” “綠 (*lù*; green),” “藍 (*lán*; blue),” “黃 (*huán*; yellow),” “紫 (*zǐ*; purple),” “橘 (*jú*; orange),” “灰 (*huī*; gray),” “咖啡 (*kāfēi*; brown),” “白 (*bái*; white),”

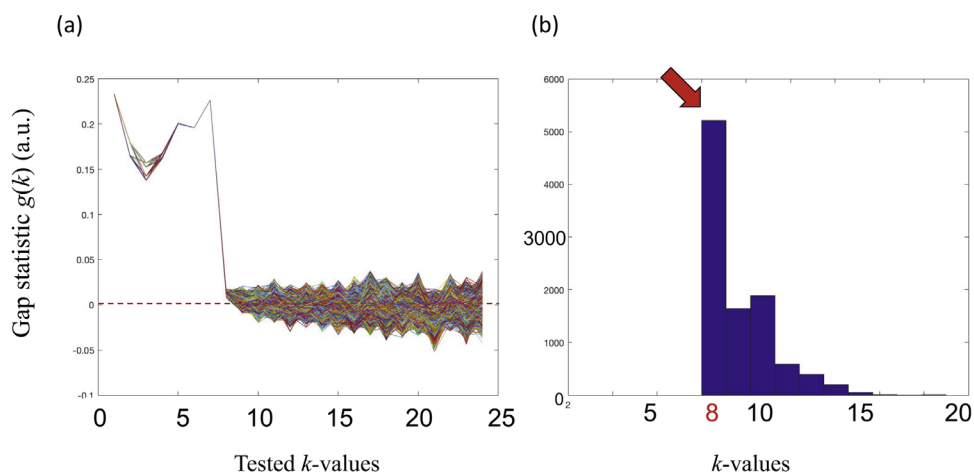


Figure 2. Gap-statistics analysis for  $k = 2$  to 24. (A) Trace of gap statistic values for 10,000 calculations. The *horizontal* and *vertical* axes represent the  $k$ -values tested and gap-statistic value for each  $k$ , respectively. (B) Histogram of cluster numbers ( $k$ ), which was the smallest number before the gap-statistic value fell from a positive value to a negative value in each calculation. The most frequent number of clusters was  $k = 8$ .

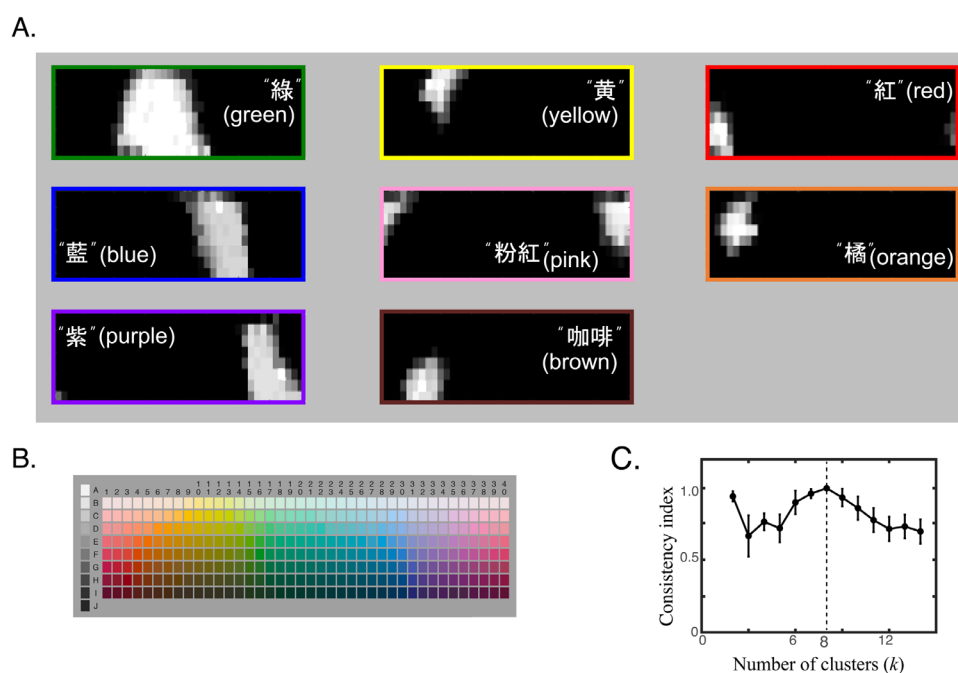


Figure 3. (A) Eight clusters derived by  $k$ -means clustering analysis of the data from 41 participants. The brightness of the cluster represents the consistency across participants. (B) Each *square* represents the arrangement of the Munsell color chip set used in the World Color Survey (Kay et al., 2011) and are corresponding to the position within each small square of panel (A). (C) The plot shows the consistency index for repeated application of  $k$ -means clustering. The *horizontal* axis shows the number of clusters, and the *vertical* axis shows the consistency index (see text for details). *Error bars* indicate 95% confidence intervals after 1000 calculations for each  $k$ . The index was maximum at  $k = 8$ .

“粉紅 (*fěnhóng*; pink),” and “黑 (*hēi*; black).” The use of “咖啡 (*kāfēi*; coffee)” for brown is a particular feature of this data set. This point will be discussed in Discussion section. Table 2 shows

that the number of participants drops sharply for words after the 11<sup>th</sup> rank. In fact, Figure 1 shows that the slope of the branch for the less popular color terms is very much steeper than Japanese

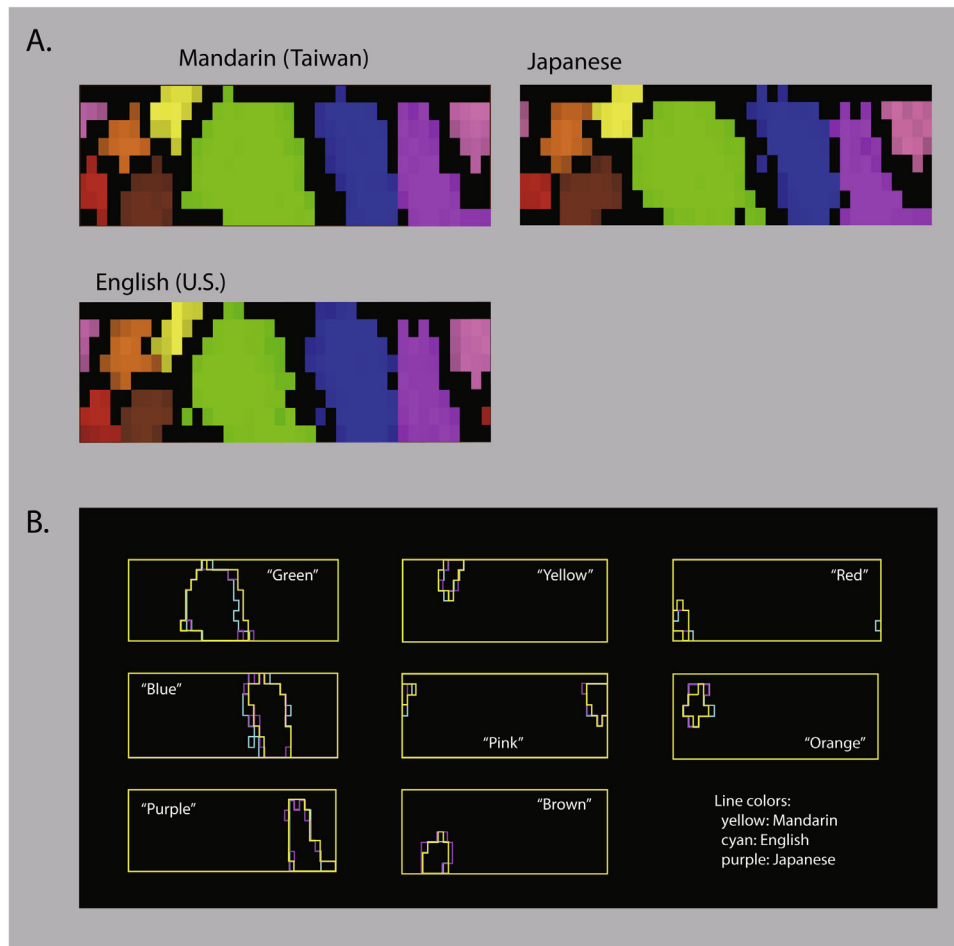


Figure 4. Comparison among Mandarin (Taiwan), English, and Japanese data sets, clustered under the same condition as the Mandarin data:  $k = 8$ . (A) The colored clusters show regions of over 80% consensus. (B) *Outlines* indicate the boundaries of clusters for corresponding categories in panel A (80% consensus) in the Mandarin (*yellow*), English (*cyan*), and Japanese (*purple*) data.

(Kuriki et al., 2017) and American English (Lindsey & Brown, 2014).

### K-means analysis

After the gap-statistics analysis was applied, the optimal number of clusters ( $k$ ) was defined as eight (Figure 2). This optimal  $k$ -value is statistically significant at the level of  $p = 0.0005$  for type-1 error, after 10,000 calculations of gap statistic value for randomly shuffled data (i.e., reference dataset prepared for Gap statistics calculation).

The results of clustering with  $k = 8$  for the Mandarin (Taiwan) data are shown in Figure 3. The label for each category was the most frequently used color term to name each color category. This label is shown with double quotation marks in the figure. Consistency index analysis endorses the reproducibility of all color categories as the highest among the numbers tested and had maximum value (1.0) at  $k = 8$ .

To maximize the benefit of using the same method as previous studies, we made comparisons with other language data sets collected following the same protocol. In an attempt to equate the number of clusters, we reapplied  $k$ -means analysis to the *chromatic* responses ( $1 \times 320$ -dimensional binary vectors) in North American English (in short, *English* in the followings; 964 vectors for 51 participants; Lindsey & Brown, 2014) and Japanese data sets (828 vectors for 57 participants; Kuriki et al., 2017) with  $k = 8$ . The resulting comparisons of clusters demonstrated over 80% consistency among participants (Figure 4A).

Figure 4B compares the outlines of area of clusters with 80% consensus for naming color chips within each participant group for Mandarin (Taiwan), English, and Japanese with yellow, cyan, and purple contour lines, respectively. The result shows a considerable similarity in the locations and area shapes of basic color categories among the three languages in general. The similarity between categories was defined using Pearson's correlation coefficient ( $\rho$ ) of vectors ( $1 \times 320$

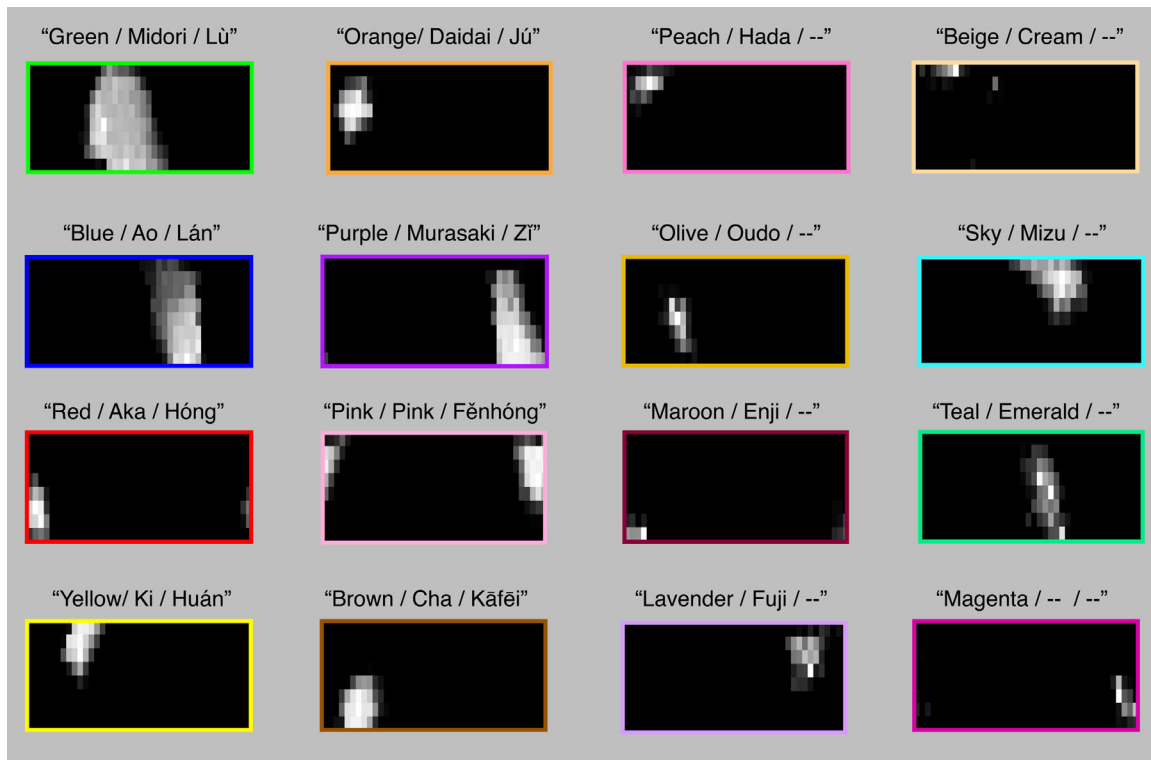


Figure 5. Sixteen optimal categories derived after pooling data from three languages. The labels of categories on the top of each panel are described in English, Japanese, and Mandarin, from left. “—” are used when naming data glossed to that color category were not found in the language. The eight categories on the left correspond to BCTs, the eight on the right to non-BCTs.

elements) for each cluster, averaged across participants, and the similarity between corresponding clusters in Mandarin and Japanese and that in Mandarin and English was quite high: on average,  $0.963 \pm 0.0187$  and  $0.945 \pm 0.0334$  (mean  $\pm$  SD), respectively. The correlation coefficients for the eight chromatic clusters were not different with statistical significance after a two-tailed  $t$ -test ( $t(7) = 1.14$ ;  $p = 0.292$ ; *NS*).

Taking a closer look at the result, we found that the border between the green and blue categories showed a slight but systematic shift between English (cyan contour) and the other two data sets (yellow and purple contours). The similarity between blue clusters of Mandarin (Taiwan) and English is relatively low (correlation coefficient  $\rho = 0.893$ ), as well as that of Japanese and English ( $\rho = 0.899$ ), while that between Mandarin (Taiwan) and Japanese is very high ( $\rho = 0.982$ ). Compared with Mandarin and Japanese, the area of the green category in English appeared slightly smaller and the blue category, slightly larger. This slight discrepancy in the green and blue areas equals one hue step in the chart (Figure 3B), equivalent to 2.5 on the Munsell hue scale. This was probably due to the use of the *teal* category in the English data set (Lindsey & Brown, 2014), which was not present in either the Mandarin (Taiwan) or Japanese data sets.

To compare in terms of the precision in communicating color information among languages with an information theory–based measure, we calculated the group mutual information (GMI), used in recent studies of color categories as a measure of communication efficiency (e.g., Conway, Ratnasingam, Jara-Ettinger, Futrell, & Gibson, 2020; Lindsey & Brown, 2019). We adopted a definition used in a previous study (Lindsey & Brown, 2015); briefly, GMI quantifies information transfer (in *bits*) within a language community, and the index increases as more terms are used more commonly among group individuals and if the variability of color name usage across individual is smaller. The GMIs from the raw data of color naming in Mandarin, Japanese, and English were 2.03 bits for Mandarin (95% confidence interval [CI] derived from 10,000 times random sampling was 1.99–2.06), 2.39 bits (95% CI: 2.35–2.43) for English, and 2.42 bits (95% CI: 2.38–2.45) for Japanese.

This result includes the effect of different numbers of color terms within each group that were  $11.6 \pm 1.05$  words for Mandarin,  $17.7 \pm 6.7$  words for Japanese, and  $21.6 \pm 7.6$  words for English, in mean  $\pm$  SD. To reduce the effect of this factor, the chromatic categories in Mandarin, English, and Japanese data sets were all glossed to eight chromatic categories by  $k$ -means in each



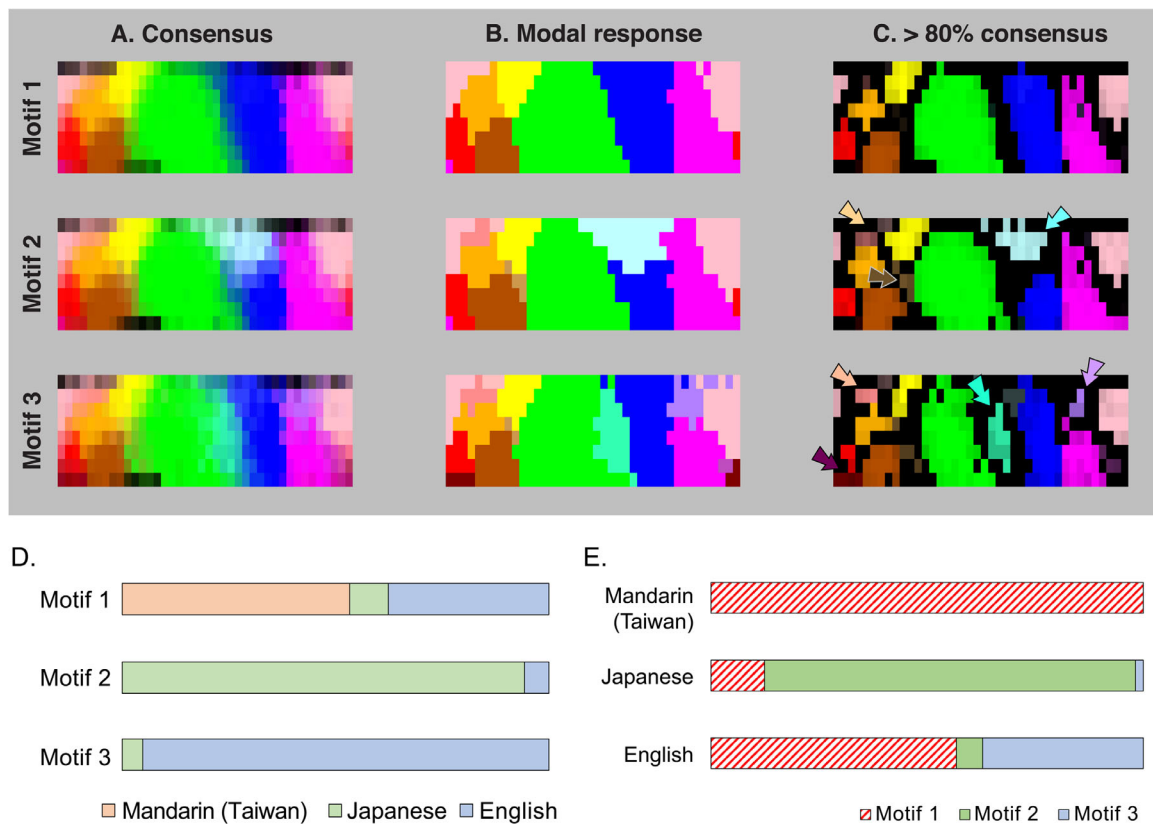


Figure 6. Motif analysis result. Rows show three types of color-naming system, i.e., motifs. (A) Consensus areas of color categories across participants. (B) Most frequent responses. (C) Areas with over 80% consensus criterion. *Arrows* show clusters that are unique in each motif. Motif 1 is a universal BCT type, Motif 2 has Green-Blue-Mizu clusters (Kuriki et al., 2017), and Motif 3 has Green-Blue-Teal clusters. (D) Fraction of speaker of each language contained in each of the 3 motifs. (E) Fraction in speakers of each language classified to each of the three motifs. See main text for precise numbers.

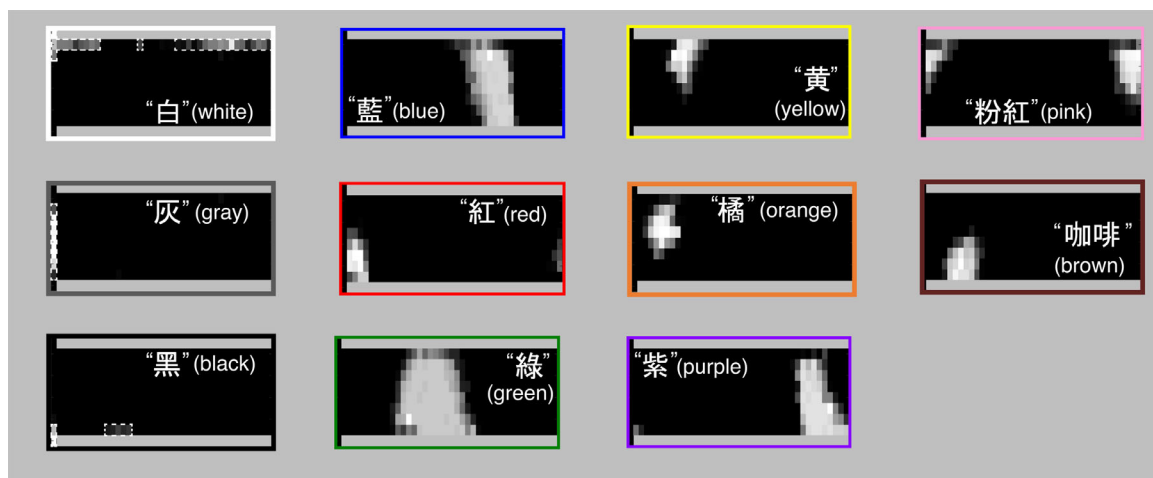


Figure 7. The analysis on the use of gray and brown terms using 330 color chips. The arrangement of colors is the same as in Figure 3B; 10 achromatic chips are now included in the leftmost column. Clusters represent the results of *k*-means analysis and category labels were chosen as the most frequently used names, indicated in each cluster. Cluster contours for achromatic terms are outlined with a broken line for the better visibility. The “灰 (*hui*; gray)” (*left middle*) shows they are confined to achromatic color chips (leftmost column). The “白 (*bai*; white)” has color chips at the top row (highest lightness) of chromatic chips and “黑 (*hei*; black)” also have four color chips at the bottom row (lowest lightness) of chromatic chips. These trends are commonly found in English (Lindsey & Brown, 2014) and Japanese (Kuriki et al., 2017) and are part of the typical distribution of white and black categories.

language group (Figure 4A) before calculating GMIs, yielding the values of 2.09 bits for Mandarin (95% CI: 2.06–2.13), 2.09 bits for English (95% CI: 2.06–2.12), and 2.18 bits for Japanese (95% CI: 2.16–2.20). The results suggest that Mandarin (Taiwan) color terms are similarly precise in communicating color information as those in Japanese and English. The remaining differences among languages may be due to individual variability in the border for color categories. For the Mandarin (Taiwan), this could be due to the presence of various synonyms to each color category (e.g., Sun & Chen, 2018).

## Motif analysis

One of the remarkable achievements of the series of studies by Lindsey and Brown (2006, 2009, 2014, 2019) was the finding of multiple color naming systems, that is, motifs. As multiple motifs occur within the same language group and sometimes commonly across language groups, we thus conducted a motif analysis by pooling data sets from the three language groups.

The analysis started by specifying a set of the optimal number of clusters (i.e., color categories) to gloss the color naming data for all three language groups. We applied *k*-means analysis with gap-statistics analysis after pooling the data for three language groups with 149 participants in total (41 Mandarin speakers, 57 Japanese speakers, and 51 English speakers) for 320 chromatic color chips. The number of chromatic responses was 2,145 in total (Mandarin 353, Japanese 828, and English 964). Gap-statistics analysis showed that the optimal *k* for the pooled data of the three languages was 16. Figure 5 shows the resulting 16 chromatic categories.

All color-naming data for the 149 participants from the three languages were glossed to these 16 color categories. Next, the fraction of color chips (out of 320) glossed to each of the 16 color categories was derived for each participant to prepare a  $1 \times 16$ -dimensional vector, where each element takes a value between 0.0 and 1.0; 149 vectors in total. According to the gap-statistics analysis of the 149 vectors, the optimal number of clusters (optimal *k*) was three.

Figure 6 shows the motifs obtained by classifying the pooled data of 16 dimensional vectors from 149 participants into three clusters. The most remarkable difference among the motifs can be seen in the classifications in the green-blue region (Figure 6C). Motif 1 used the eight chromatic categories of universal BCTs only; Motif 2, Green-Blue-Mizu categorization (Kuriki et al., 2017); and Motif 3, Green-Blue-Teal categorization patterns. The population of language speakers in each motif were as follows: Motif 1, 41 Mandarin, 7 Japanese, and 29 English-speaking participants; Motif 2, 0 Mandarin, 49 Japanese, and

3 English; and Motif 3: 0 Mandarin, 1 Japanese, and 19 English (Figure 6D). Arrows in Figure 6C indicate categories that are unique in each motif: *mizu* (light blue in Japanese), *oudo* (tan), and *hada* (skin) categories in Motif 2, and lavender, teal, peach, and maroon categories in Motif 3.

This analysis indicated that Motif 1 with eight chromatic categories best represented current Mandarin (Taiwan) participants' color naming structure, although there was no Mandarin (Taiwan) participant classified as showing Motif 2 or 3. Such a high degree of congruency (100% Motif 1) in the usage of BCTs was not seen in Japanese and English participants (Figure 6E). The majority (86.0%) of the Japanese participants were classified as showing Motif 2 with Green-Blue-Mizu categories, whereas the U.S. participants were mostly split between the two motifs: 56.8% and 37.3% of participants, respectively, were classified as showing Motif 1 and Motif 3, which used Green-Blue-Teal categories.

## Gray and brown categories in Taiwanese Mandarin Chinese

One of the unsettled issues regarding BCT of Mandarin Chinese is whether gray is a BCT or secondary/tertiary color. Berlin and Kay's initial survey (1968) claimed gray was a BCT, but they rejected this later due to its conflict with their theory of the chronological emergence order of BCTs. A few decades later, the use of the gray (“灰 (*huī*)”) and brown terms (“咖啡 (*kāfēi*)”, “褐 (*hé*)”, and “棕 (*zōng*)”) remains controversial in the context of the “wild-card” theory (Gao & Sutrop, 2014; Greenfield, 1986), in which the color chips that are ambiguous for participants to name are referred to by gray or brown. We made an attempt to address this issue by applying another *k*-means clustering analysis to the data of Mandarin (Taiwan). If the gray and brown color terms were used to name color chips that were difficult to name a color, the clustering of color chips named by gray or brown would be found in areas different from those in other languages like English (Lindsey & Brown, 2014) or Japanese (Kuriki et al., 2017).

To complement the analysis, we performed an additional *k*-means clustering analysis by including achromatic color chips and responses with achromatic color terms, thus including all 330 color chips for the total 487 responses from Mandarin speakers in Taiwan. We applied a *k*-means clustering method using MATLAB genuine function (*evalclusters* function). By default, this function uses the principal component analysis (PCA) of actual data for the derivation of the reference data set, which samples random data while

taking the variability of the data (along its principal components) into account. This function also gives the optimal number for  $k$  by the gap-statistics analysis (Tibshirani et al., 2001), by choosing a “gap” option for the optimization criterion.

The results of  $k$ -means analysis for the optimal number of cluster:  $k = 11$  is shown in Figure 7. For the “灰 (*huī*; gray)” category, the color chips were obviously confined to the region of achromatic color chips (the left-most column of the panel) at the medium lightness level and did not distribute elsewhere. Similarly, the brown cluster (labeled “咖啡 (*kāfēi*; brown)” in Figure 7) distributed in an area that is commonly named brown in English and Japanese; this was confirmed by Pearson’s correlation coefficients ( $\rho$ ) between the vectors of Mandarin (Taiwan) and English ( $\rho = 0.981$ ) and between Mandarin (Taiwan) and Japanese ( $\rho = 0.932$ ). Therefore the use of gray and brown terms in the present result did not show that these terms were used as “wild cards” (Gao & Sutrop, 2014; Greenfeld, 1986) to name color chips that the participants found too ambiguous to name. Furthermore, compared with the color chips chosen for traditional brown terms in Sun & Chen (2018), the brown cluster in our study labeled “咖啡 *kāfēi*” yields high similarity when evaluated by Pearson’s correlation coefficients ( $\rho$ ) between the vectors ( $1 \times 320$  elements) in data for Sun & Chen (2018; by courtesy of authors) and our data; they were  $\rho = 0.871$  and  $\rho = 0.857$  for 棕 (*zōng*) and 褐 (*hé*), respectively.

## Evaluation of overlaps between categories

As mentioned in the Methods section, the degree of overlap between “粉紅 (*fěnhóng*; pink)” and “紅 (*hóng*; red)” was evaluated to justify the use of “粉紅 (*fěnhóng*)” for pink as an independent term from “紅 (*hóng*; red).”

Figure 8A shows the histogram of overlap ratios after 10,000 repetitions of calculations of the overlaps between “粉紅 (*fěnhóng*)” versus “紅 (*hóng*)” (median: 31%). Figure 8A also shows the histogram of overlap ratios in Japanese for *pink* (English loanword) versus *aka* (red) (median: 43%) and in English for pink versus red (median: 32%) for comparison. It shows that the “粉紅 (*fěnhóng*; pink)” and “紅 (*hóng*; red)” are more distinct than *aka* (red) and *pink* in Japanese (Kuriki et al., 2017), although “粉紅 (*fěnhóng*; pink)” appears as a variation of “紅 (*hóng*; red)” with a modifier “粉 (*fěn*; pale).” Figure 8B shows the overlap ratios for categories labeled with blue and green in Mandarin, English, and Japanese. It is evident that “粉紅 (*fěnhóng*; pink)” versus “紅 (*hóng*; red)” in Mandarin are as distinct as “藍 (*lán*; blue)” versus “綠 (*lù*; green).” Overlapping ratios between color names in Mandarin for other pair

of categories are also shown in supplemental materials (Supplementary Figure S2).

## Discussion

The results derived from a series of analyses, including  $k$ -means clustering, overlapping analysis, GMI analysis, and motif analysis, finely depict the use of modern Mandarin Chinese color terms based on the minimum constraint of monolexemic principle in the color categories against the World Color Survey chart. We also paid special attention to verify several controversial categories, including pink, brown, and gray. Current findings lead to discussion points of three perspectives: a reconsideration of the definition of BCTs, the language characteristics of current Mandarin (Taiwan) BCTs data, and the cross-languages comparison among unrelated languages.

### Reconsideration of BCTs definition

The descriptions about BCTs commonly emphasize certain linguistic features including monolexemy, hypernyms, context-independent, non-foreign loan words and object names (Berlin & Kay, 1969; Kay, 1978; Biggam, 2012), and numerous later studies adopted them as certifying rules to recognize BCTs. However, there has been studies concerning the appropriateness of applying these criteria to worldwide languages, particularly to non-English-like languages (e.g., Levinson, 2000; Biggam 2012; Brown et al., 2016; Witzel, 2019). In the issue of Mandarin Chinese BCTs, those criteria were sometimes decisive in determining the development status of color naming. Studies that rigorously used the criteria of linguistic features tended to state the numbers of Mandarin Chinese BCTs to be fewer than 11 (Wu, 2011; Gao, 2014; Gao & Sutrop, 2014), whereas those conducting color naming task have claimed 11 color categories mainly by descriptive naming frequency (Lu, 1997; Lin et al., 2001b; Hsieh & Chen, 2011). The current study made a breakthrough in justifying each of the 11 categories by sophisticated statistical means. Particularly, the pink term 粉紅 (*fěnhóng*) has been rejected as a BCT (Lu, 1997; Wu, 2011; Gao, 2014; Sun & Chen, 2018) due to its compound form rather than lacking universality in practical use. Figure 8 illustrates that pink and red are two mutually exclusive categories, although the linguistic label of the pink category contains a red BCT. This result indicates that some BCTs certified with empirical, behavioral methods could directly conflict with classic criteria emphasizing linguistic features. Consequently, applying the criteria advanced with BCT studies of a few Western languages could lead

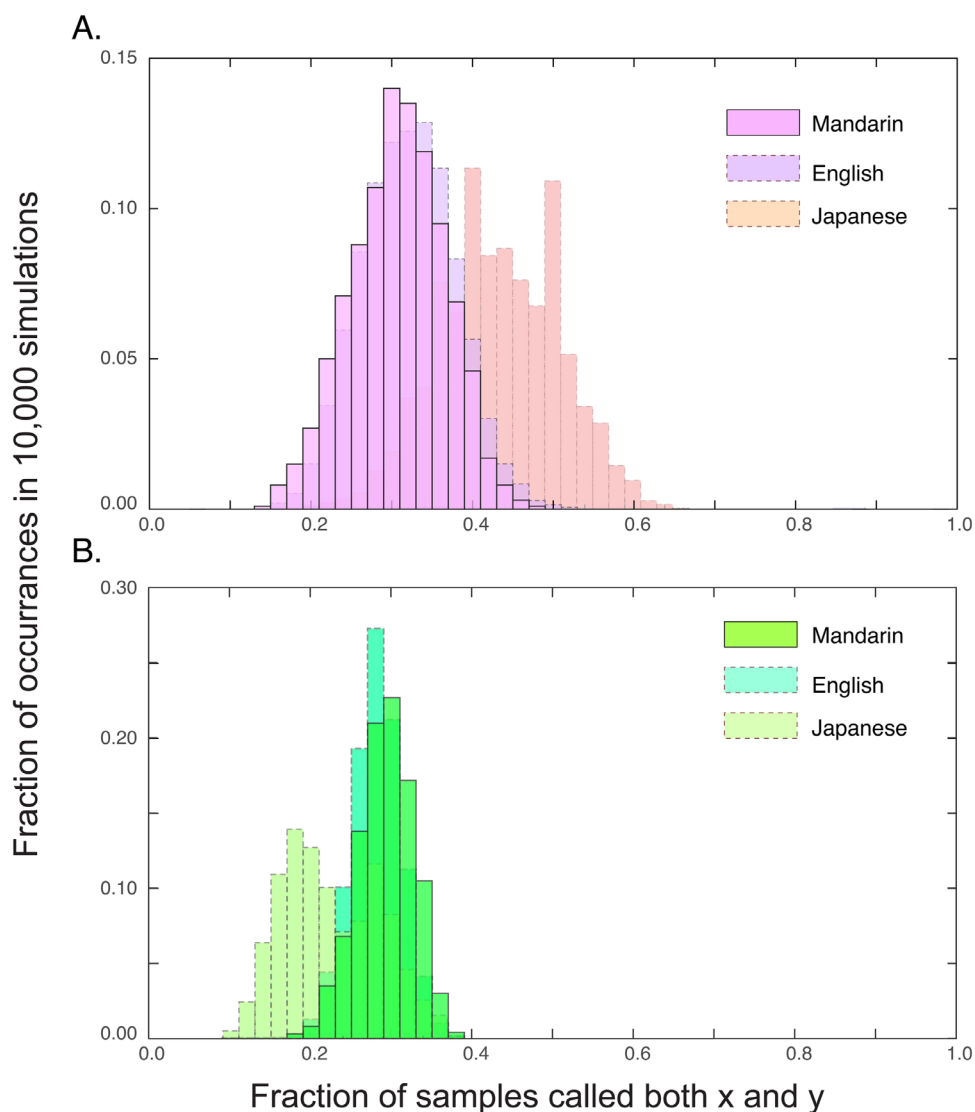


Figure 8. Analysis of category overlaps. (A) Fraction of occurrence of outcomes during 10,000-time random sampling, in which a color chip that was called  $x =$  “粉紅 (*fěnhóng*; pink)” by at least one of the 10 participants was also called  $y =$  “紅 (*hóng*; red)” by at least one other subject. Histograms for the overlap ratio for  $x =$  red,  $y =$  pink in English and  $x =$  aka,  $y =$  pink in Japanese (from Figure 8 of Kuriki et al., 2017) are shown by the dotted lines. (B) The result of same analysis on blue and green categories ( $x =$  “藍 (*lán*; blue),”  $y =$  “綠 (*lù*; green)” for Mandarin (Taiwan);  $x =$  ao,  $y =$  midori in Japanese) for three language groups.

to misestimating the number of BCTs in a different language family. Therefore our findings suggest the necessity of promoting the importance of behavioral metrics to the BCT criterion of saliency, such as consistency and elicibility, while investigating basic color categories and BCTs in diverse languages.

### Notable language characteristics in current Mandarin (Taiwan) BCTs

One of the notable characteristics in the current dataset is less diverse BCTs when compared with the

previous studies. The number of elicited color terms in total (see Table 2) is fewer than that in Japanese and English studies (Lindsey & Brown 2014; Kuriki et al., 2017). The lack of variety also resulted in the lower optimal clustering number of 11 (eight chromatic clusters plus three achromatic clusters) in current data (Figures 3 and 7). In contrast, Japanese and English optimal numbers were 16 and 17, respectively (Kuriki, 2017; Lindsey and Brown, 2014). Also, the motif analysis suggests the current Mandarin (Taiwan) data appear rather uniform as all participants were grouped into a same motif type (Figure 6E). In addition, the dataset also missed traditional, historical color terms mentioned in studies involving the corpus method

(Wu, 2011; Sun & Chen, 2018) or in the earlier study (Lu, 1997). We speculated that the monolexemic constraint applied during the collecting procedure could be one of the reasons behind the reduction in the variety of observed color terms. The Mandarin language, in general, does not allow for strict evaluation of the monolexemic criterion because most color words in Mandarin are composed of at least two characters, which represent single lexemes in Mandarin. In most contexts of real use, the expression of color is usually added with a suffix “色 sè” meaning color; for example, red is expressed as 紅色 (hóng-sè). There are plentiful Mandarin color terms named with various objects that is similar to Japanese color terms “空 (sora; sky)” and “水 (mizu; water)” (Kuriki, 2017) or English color terms violet and salmon (Lindsey & Brown, 2014). These terms also rarely present with single object name but always present with a suffix of common BCTs, e.g., “米白 mǐ bái (rice-white),” “水藍 shuǐ lán (water-blue),” etc. The variety of color naming in modern Mandarin Chinese exhibits better in compound color terms instead of single-worded color terms.

Despite that the monolexemic restriction limits the word choices from the pool of modern color terms, the participants could still have had plenty of choices from traditional, monolexemic color terms like those listed in Sun & Chen (2018) and Wu’s corpus studies (2011, 2014). Sun and Chen’s study (2018) demonstrated that their participants of similar backgrounds with ours (young college students in Taiwan) were capable of selecting proper color chips to match historical color terms. They identified multiple color terms to constitute the “color family” (Sun and Chen, 2018); for example, the concept of green family was composed of five terms with corresponding color chips, 碧 bì, 菜 cài, 翠 cuì, 綠 lù, and 青 qīng, while red family was six: 赤 chì, 丹 dān, 紅 hóng, 血 xuè, 赭 zhě, and 朱 zhū. The Pearson’s correlation coefficients ( $\rho$ ) between the vectors (each  $1 \times 320$  dimension) for clusters in the present study and some of traditional color terms in Sun and Chen’s are reasonably high, for example,  $\rho = 0.906$  for pink terms, 粉紅 fěnhóng versus 櫻 yīng;  $\rho = 0.871$  for brown terms, 咖啡 kāfēi versus 棕 zōng, indicating that color synonyms within a color category could be common in Mandarin Chinese. Curiously, there were few participants in the present study who used traditional alternative color terms. For example, the orange term 橘 jú ( $N = 41$ , 100%) significantly prevailed over another orange term 橙 chéng ( $N = 2$ ), although 橙 chéng was reported to be an orange BCT earlier (Lu, 1997). The foreign loan word 咖啡 kāfēi ( $N = 41$ , 100%) prevailed over traditional brown terms 棕 zōng ( $N = 10$ ) and 褐 hé ( $N = 2$ ), that were considered promising as BCTs (e.g., Lu, 1997; Lin et al., 2001). We speculate that generational or regional differences or both could be one of the causes to explain such phenomena. Perhaps to the younger residents in internationalized

area (Taipei city) like our participants, color terms like 咖啡 (kāfēi) are just trendy colloquial expression with plain-spoken, easily-recalled, and foreign-cultural features. They might have spontaneously preferred using it even though they were taught many classic choices of synonyms or hypernyms referring to brown, repeatedly in school textbooks.

On the other hand, some BCTs found in the present study are consistent with previous studies that have been reported constantly across contemporary and historical corpus studies. Specifically, 白 bái (white) and 黃 huáng (yellow) have been ancient BCTs for more than 30 centuries, whereas 黑 hēi (black), 紅 hóng (red), 綠 lù (green), 藍 lán (blue), and 紫 zǐ (purple) have been used for more than 10 centuries (Wu, 2011, 2014; Gao, 2014). All these were also reported as modern Mandarin Chinese BCTs in agreement across color naming studies (e.g., Lu, 1997; Hsieh & Chen, 2011; Gao, 2014; Gao & Sutrop, 2014). It seems that the level of basicness reflects not only the naming consistency in lab-based behavioral research but also the frequent practical use in real-world communication. Thus we did a supplemental corpus statistic to compare the word frequency of our data and other commonly reported color terms in the balanced linguistic corpus: *Accumulated Word Frequency in Modern Chinese Corpus*, established by the Institute of Linguistics in Academia Sinica Taiwan. It is a *balanced* linguistic corpus that targets 10 million words from spoken and written materials of various themes evenly sampled between 1981 to 2007 (for details, see Chen, Huang, Chang, & Hsu, 1996; Wu, Jin, Zhang, & Yu, 2006). We searched 17 color terms, including modern BCTs and several synonyms/hypernyms, in this database.

The colored thick bars in Figure 9 shows the word frequency of each 17 color terms in modern era (from 1981–2007) against the gray bars presenting that of early eras (10th–19th CE) and white bars presenting those in ancient eras (third BCE to first CE). Such cross-era comparison reveals significant changes in the prevalence of color terms with time. For example, 青 qīng was one of the frequently used color terms during the early era, but it appears less popular in the modern era. The loanword 咖啡 kāfēi has not appeared until the modern era but it is now more popular than 棕 zōng and 褐 hé in our color naming experiment. We analyzed the correlations between the numbers of common color terms present both in the linguistic corpus of modern era and in our data listed in Table 2, and found them positively and significantly correlated (Pearson’s correlation coefficient:  $\rho = 0.52$ ,  $p = 0.03$ ). However, the correlation between the numbers of current common color terms and those in the corpus of the early era was relatively insignificant ( $\rho = 0.32$ ;  $p = 0.14$ ). The comparison of our data collected in the lab and corpus-based data shows converging tendency in the use of common Mandarin Chinese color terms,

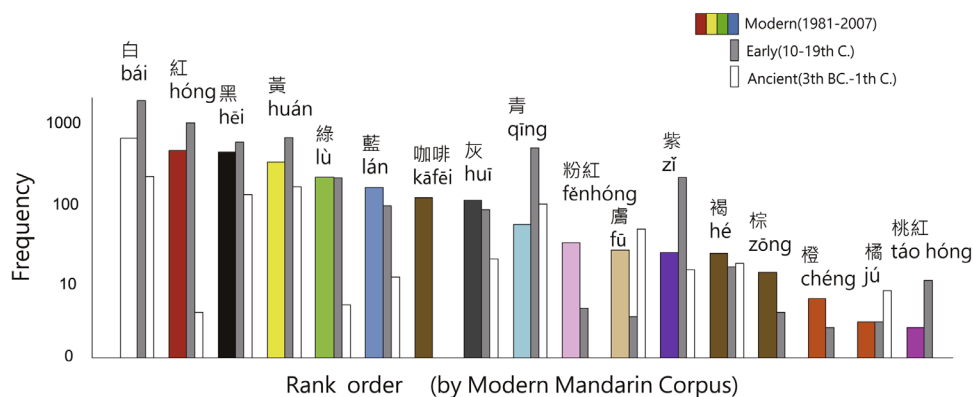


Figure 9. Frequency rank of color terms in Academia Sinica Balanced Corpus of Modern Chinese (*colored bars*), Academia Sinica Tagged Corpus of Early Mandarin Chinese (*gray bars*), and Academia Sinica Ancient Chinese Corpus (*white bars*). The *X-axis* presents the frequency rank of color terms based on the results of the modern corpus, whereas the *Y-axis* (log scale) presents the count of each color term.

which improve the fidelity of current findings in Mandarin BCTs.

Among the above discussion on Mandarin Chinese color terms across eras and studies, 青 *qīng* might be another color term worth mentioning because it has undergone significant transitions in terms of popularity and meanings. It initially used to describe a wide range of greenish to bluish shades in ancient times, then turned to be a grue term from the Tang dynasty (7th to 10th CE.; Wu, 2011; Gao, 2014). Its contemporary meaning could vary depending on the regions of speakers. In Mainland China, *qīng* specifically refers to cyan or teal, the greenish-blue to blue-green color, and it is known as one of the terms for naming rainbow colors in between green and blue. However, *qīng* is a less common color term for some Mandarin speakers in Taiwan and Hong Kong, where the elementary education does not include *qīng* as a rainbow color. Perhaps that was the reason many adults rarely used it in daily communication, as well as that observed in the current study. Contrasting with definite meaning of *qīng* known in Mainland China, Sun & Chen (2018) demonstrated that younger Taiwan residents recognized *qīng* as a vague “grue” term or seemingly a synonym of green because the corresponding color chips matched *qīng* mainly deployed in green area and extending to blue. It exhibited another kind of evolution after *qīng* was adopted to Japanese. The case of *qīng* is an excellent example to highlight the essence of living languages as an evolving cultural phenomenon affected by multifaceted factors.

## Comparison with nonrelated languages

One of the purposes of choosing the present method is to compare the current data with the English study

(Lindsey & Brow, 2014) and Japanese study (Kuriki et al., 2017). English, Japanese, and Mandarin are classified to three unrelated language families, Indo-European, Altaic, Sino-Tibetan, respectively. However, the intensive social and economic interflows seem to increase the cultural homogeneity between speakers of the three languages, and such tendency is also reflected in our analysis of cross-languages comparison. Although there were cautions that the World Color Survey chart is composed of maximum Chroma of each hue and that could have led to spurious similarities (e.g., Witzel, 2019), it is notable that the extremely high similarity in the manner of partitioning World Color Survey charts among the three linguistically unrelated languages, as shown in Figure 4 in which cluster number was set at  $k = 8$ . The correlations between the categories of the three languages explain more than 88% of the variance, leaving little room for variation due to cultural differences.

The particularly high similarity between Mandarin and Japanese could be due to long-term, substantial cultural exchanges. Most kanji characters in Japanese and their meanings were imported from China. For example, 青 *ao* means blue and sometimes green in the Japanese dictionary. This is possibly because the letter 青 was borrowed from Chinese during the *grue era* of Japan (probably fifth century CE, before the separation of blue and green in ninth to tenth century CE; Stanlaw, 2010; Kuriki et al., 2017), and it was used to write the native Japanese term *ao*, which meant *grue* in that era. Blue in modern Mandarin Chinese is “藍 (*lán*),” whereas “藍 (*ai*)” in Japanese means indigo (deep/dark blue) and is the name of a bluish dye for textiles (indigo) in both languages. In recent Japanese, a distinction between light blue, 水 (*mizu* in JP; *shuǐ* in MC; both literally ‘water’) or 空 (*sora* in Japanese; *kōng* in Mandarin; both literally “sky”),

from darker blue 青 ao has been established in the past 30 years (Kuriki et al., 2017; Uchikawa & Boynton, 1987; Kuriki, 2019), although this distinction does not appear in our data for Mandarin Chinese speakers in Taiwan. This clearly illustrates that changes in the clusters could take place independently in each cultural or linguistic group, whereas each word shares its original meaning among both languages in other cases (Kay, 1978; Biggam, 2012; Witzel, 2019). Thus these two language groups partially use the same characters to name similar color categories, but the differences in cultural background and history after being imported to Japan have established unique modifications in each language group.

A slight systematic difference was observed between the two groups across the Pacific Ocean at the border between the blue and green categories (Figure 4B). The green/blue border in English is slightly shifted toward yellow by 2.5 steps in Munsell hue. This may be due to the presence of the teal category located in the medium lightness level between the blue and green categories (Motif 3 in Figure 6; also see Lindsey & Brown, 2014).

## Conclusion

We investigated Mandarin Chinese basic color terms in Taiwanese participants by using a quantitative *k*-means clustering approach, which has not previously been implemented on the issue. We found that Mandarin basic color terms refer to eleven basic color categories. The GMI analysis suggested that Mandarin color terms convey largely the same information about color as do English and Japanese color terms. These findings show quite clearly that fundamental differences in history, culture, and etymology do not impose fundamental differences in the modern use of color terms. For orange, pink, and brown categories, we supplemented the comparison with word frequency data across eras to demonstrate the possibility of using cultural or generational differences to explain the discrepancy among color naming data between studies. The contradictions between behavioral measures and certain linguistic criteria of BCTs, observed in the categories like pink, appeal to refine the definition of BCTs that are applicable to a wider variety of language. The additional cross-era comparison suggested that some color terms have been more stably used than others. The particular case of 青 seem to vary in its prevalence and meanings across eras, regions, and even languages (Mandarin and Japanese), which implies that the stability of meaning of the color terms could be another index to evaluate the basicness.

*Keywords:* color categories, basic color terms, mandarin chinese, clustering analysis

## Acknowledgments

The authors thank Delwin T. Lindsey and Angela M. Brown for providing the data of their English study and their help with the analysis methods. We also thank Vincent C. Sun and Chien-Chung Chen for providing their data.

Supported by a funding program of RIEC, Tohoku University (Nationwide Cooperative Research Program) for international collaborations H26/A26 and H29/A19 “Similarity and dissimilarity of color categories across language/cultural difference and among individuals.” Also supported by the Taiwan Ministry of Science and Technology No. 105-2410-H-034-037, 106-2410-H-034 -034 -MY2, 109-2410-H-034 -010 to TH, and Japan Society for the Promotion of Science (JSPS) KAKENHI Nos. JP15H03460, JP18H04995, and JP20H00597 to IK.

Commercial relationships: none.

Corresponding author: Ichiro Kuriki.

Email: [ikuriki@riec.tohoku.ac.jp](mailto:ikuriki@riec.tohoku.ac.jp).

Address: Research Institute of Electrical Communication, Tohoku University, 2-1-1, Katahira, Aoba, Sendai 980-8577, Japan.

## References

- Biggam, C. P. (2012). *The Semantics of Colour: A historical approach*. Cambridge: Cambridge University Press.
- Brown, R. W., & Lenneberg, E. H. (1954). A study in language and cognition. *Journal of Abnormal and Social Psychology*, 49(3), 454–462.
- Brown, A. M., Isse, A., & Lindsey, D. T. (2016). The color lexicon of the Somali language. *Journal of vision*, 16(5), 14–14.
- Berlin, B., & Kay, P. (1969). *Basic color terms: Their universality and evolution*. Berkeley: University of California Press.
- Chen, K. J., Huang, C. R., Chang, L. P., & Hsu, H. L. (1996). Sinica corpus: Design methodology for balanced corpora. *Proceedings of the 11th Pacific Asia Conference on Language, Information and Computation* (pp. 167–176).
- Conway, B. R., Ratnasingam, S., Jara-Ettinger, J., Futrell, R., & Gibson, E. (2020). Communication efficiency of color naming across languages provides a new framework for the evolution of color terms. *Cognition*, 195, 104086. <https://doi.org/10.1016/j.cognition.2019.104086>.

- Gao, J., & Sutrop, U. (2014). The basic color terms of Mandarin Chinese: A theory-driven experimental study. *Studies in Language*, 38(2), 335–359, doi: [10.1075/sl.38.2.03gao](https://doi.org/10.1075/sl.38.2.03gao).
- Gao, J. (2014). Basic Color Terms in Chinese: Studies after the Evolutionary Theory of Basic Color Terms (Doctoral dissertation).
- Greenfield, P. J. (1986). What is gray, brown, pink, and sometimes purple: The range of “wild-card” color terms. *American Anthropologist*, 88(4), 908–916.
- C. L. Hardin, C. L. Hardin, & L. Maffi (Eds.). (1997). *Color categories in thought and language*. Cambridge: Cambridge University Press.
- Heider, E. R. (1972). Universals in color naming and memory. *Journal of Experimental Psychology*, 93(1), 10–20, [http://doi:10.1037/h0032606](https://doi.org/10.1037/h0032606).
- Hsieh, T. J., & Chen, I. P. (2011). Categorical formation of Mandarin color terms at different luminance levels. *Color Research & Application*, 36(6), 449–461, [http://doi:10.1002/col.20638](https://doi.org/10.1002/col.20638).
- Ishihara, S. (1918). Tests for color blindness. *American Journal of Ophthalmology*, 1(5), 376.
- Jraissati, Y. (2014). Proving universalism wrong does not prove relativism right: Considerations on the ongoing color categorization debate. *Philosophical Psychology*, 27(3), 401–424, doi: [10.1080/09515089.2012.733815](https://doi.org/10.1080/09515089.2012.733815).
- Kay, P., & McDaniell, C. K. (1978). The linguistic significance of the meanings of basic color terms. *Language*, 54(3), 610–646, [http://doi:10.2307/412789](https://doi.org/10.2307/412789).
- Kay, P., & Regier, T. (2003). Resolving the question of color naming universals. *Proceedings of the National Academy of Sciences*, 100(15), 9085–9089, [http://doi:10.1073/pnas.1532837100](https://doi.org/10.1073/pnas.1532837100).
- Kay, P., Berlin, B., Maffi, L., Merrifield, W. R., & Cook, R., (2011). *The World Color Survey*. Stanford: Stanford University.
- Kuriki, I. (2019). Emergence and separation of color categories: An NIRS study in prelingual infants and a *k*-means analysis on Japanese color-naming data. *Current Opinion in Behavioral Sciences*, 30, 21–27, [http://doi:10.1016/j.cobeha.2019.04.012](https://doi.org/10.1016/j.cobeha.2019.04.012).
- Kuriki, I., Lange, R., Muto, Y., Brown, A. M., Fukuda, K., Tokunaga, R., . . . Shioiri, S (2017). The modern Japanese color lexicon. *Journal of Vision*, 17(3), 1, [http://doi:10.1167/17.3.1](https://doi.org/10.1167/17.3.1).
- Levinson, S. C. (2000). Yéfi Dnye and the theory of basic color terms. *Journal of Linguistic Anthropology*, 10(1), 3–55.
- Lin, H., Luo, M. R., MacDonald, L. W., & Tarrant, A.W. S. (2001a). A cross-cultural colour-naming study: Part I: Using an Unconstrained Method. *Color Research and Application*, 26(2), 40–60, [http://doi:10.1002/1520-6378\(200102\)26:13.0.CO;2-X](https://doi.org/10.1002/1520-6378(200102)26:13.0.CO;2-X).
- Lin, H., Luo, M. R., MacDonald, L. W., & Tarrant, A.W. S. (2001b). A cross-cultural colour-naming study: Part II – Using a Constrained Method. *Color Research and Application*, 26(3), 193–208, [http://doi:10.1002/col.1017](https://doi.org/10.1002/col.1017).
- Lindsey, D. T., & Brown, A. M. (2006). Universality of color names. *Proceedings of the National Academy of Sciences*, 103(44), 16608–16613, [http://doi:10.1073/pnas.0607708103](https://doi.org/10.1073/pnas.0607708103).
- Lindsey, D. T., & Brown, A. M. (2009). World Color Survey color naming reveals universal motifs and their within-language diversity. *Proceedings of the National Academy of Sciences*, 106(47), 19785–19790, [http://doi:10.1073/pnas.0910981106](https://doi.org/10.1073/pnas.0910981106).
- Lindsey, D. T., & Brown, A. M. (2014). The color lexicon of American English. *Journal of vision*, 14(2), 17, [http://doi:10.1167/14.2.17](https://doi.org/10.1167/14.2.17).
- Lindsey, D. T., Brown, A. M., Brainard, D. H., & Apicella, C. L. (2015). Hunter-gatherer color naming provides new insight into the evolution of color terms. *Current Biology*, 25(18), 2441–2446, [http://doi:10.1016/j.cub.2015.08.006](https://doi.org/10.1016/j.cub.2015.08.006).
- Lindsey, D. T., & Brown, A. M. (2019). Recent progress in understanding the origins of color universals in language. *Current Opinion in Behavioral Sciences*, 30, 122–129, [http://doi:10.1016/j.cobeha.2019.05.007](https://doi.org/10.1016/j.cobeha.2019.05.007).
- Lu, C. F. (1997). Basic Mandarin color terms. *Color Research and Application*, 22(1), 4–10, [http://doi:10.1002/\(SICI\)1520-6378\(199702\)22:1<::AID-COL3>3.0.CO;2-Z](https://doi.org/10.1002/(SICI)1520-6378(199702)22:1<::AID-COL3>3.0.CO;2-Z).
- Mylonas, D., & Griffin, L. D. (2020). Coherence of achromatic, primary and basic classes of colour categories. *Vision Research*, 175, 14–22.
- Saunders, B. (2000). Revisiting Basic Color Terms. *Journal of the Royal Anthropological Institute*, 6(1), 81–99.
- Siuda-Krzywicka, K., Witzel, C., Bartolomeo, P., & Cohen, L. (2020). Color naming and categorization depend on distinct functional brain networks. *bioRxiv*, [https://doi:10.1101/2020.04.13.038836](https://doi.org/10.1101/2020.04.13.038836).
- Stanlaw, J. (2010). Language, contact, and vantages: Fifteen hundred years of Japanese color terms. *Language Sciences*, 32(2), 196–224, [http://doi:10.1016/j.langsci.2009.10.005](https://doi.org/10.1016/j.langsci.2009.10.005).
- Sun, V. C., & Chen, C. C. (2018). Basic color categories and Mandarin Chinese color terms. *PLoS One*, 13(11), e0206699, [http://doi:10.1371/journal.pone.0206699](https://doi.org/10.1371/journal.pone.0206699).
- Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of clusters in a data set via



- the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(2), 411–423, <http://doi:10.1111/1467-9868.00293>.
- Uchikawa, K., & Boynton, R. M. (1987). Categorical color perception of Japanese observers: Comparison with that of Americans. *Vision Research*, 27(10), 1825–1833, [http://doi:10.1016/0042-6989\(87\)90111-8](http://doi:10.1016/0042-6989(87)90111-8).
- Whorf, B.L. (1956). *Language, Thought and Reality: Selected Writings of Benjamin Lee Whorf*. Cambridge: MIT Press.
- Witzel, C. (2019). Misconceptions about colour categories. *Review of Philosophy and Psychology*, 10(3), 499–540, <http://doi:10.1007/s13164-018-0404-5>.
- Wu, J. (2011). The evolution of basic color terms in Chinese. *Journal of Chinese linguistics*, 39(1), 76–122.
- Wu, Y., Jin, P., Zhang, Y., & Yu, S. (2006). A Chinese corpus with word sense annotation. In Y. Matsumoto, R.W. Sproat, K. F. Wong, & M. Zhang (Eds.), *Computer Processing of Oriental Languages. Beyond the Orient: The Research Challenges Ahead* (pp. 414–421). Heidelberg, Berlin: Springer, [https://doi.org/10.1007/11940098\\_43](https://doi.org/10.1007/11940098_43).