

# The color lexicon of American English

**Delwin T. Lindsey**

Department of Psychology, Ohio State University,  
Mansfield, OH, USA  
College of Optometry, Ohio State University,  
Columbus, OH, USA



**Angela M. Brown**

College of Optometry, Ohio State University,  
Columbus, OH, USA



**This article describes color naming by 51 American English-speaking informants. A free-naming task produced 122 monolexic color terms, with which informants named the 330 Munsell samples from the World Color Survey. Cluster analysis consolidated those terms into a glossary of 20 named color categories: the 11 Basic Color Term (BCT) categories of Berlin and Kay (1969, p. 2) plus nine nonbasic chromatic categories. The glossed data revealed two color-naming motifs: the green–blue motif of the World Color Survey and a novel green–teal–blue motif, which featured peach, teal, lavender, and maroon as high-consensus terms. Women used more terms than men, and more women expressed the novel motif. Under a constrained-naming protocol, informants supplied BCTs for the color samples previously given nonbasic terms. Most of the glossed nonbasic terms from the free-naming task named low-consensus colors located at the BCT boundaries revealed by the constrained-naming task. This study provides evidence for continuing evolution of the color lexicon of American English, and provides insight into the processes governing this evolution.**

## Introduction

Humans can discriminate on the order of  $10^6$  different colors, many more colors than any individual can name reliably. These colors fall into a much smaller number of categories that speakers in a language community can name and can use among themselves to communicate about color. People around the world differ greatly in the number of these named color categories. However, despite more than 150 years of research, several unresolved issues persist regarding cross-cultural differences in color lexicon. One of these issues concerns how best to characterize the relative importance of terms in a language's color lexicon. Another is how to compare and contrast color lexicons across the world's 7,000 living

languages. The third issue concerns the evolution of a language's color lexicon as color categories change. The present study seeks to address some of these issues by analyzing the color-naming behavior of a group of native English-speaking informants drawn from the relatively culturally homogeneous population of Ohio State University faculty, staff, and students.

## Berlin and Kay

The context for this work is the classic theoretical analysis of cross-cultural differences in color categories by Berlin and Kay (1969). On the basis of their study of 98 world languages, these authors advanced two conjectures about the differences they observed. Their first conjecture was that there is a limited set of *basic color terms* (BCTs) in most languages, which are distinct from other color terms that an individual might use to name colors. According to this first conjecture, the colors in the lexicon of each language are a subset drawn from a universal set of 11 color categories, which are closely related to the BCTs of English and other languages spoken in technologically advanced societies. Berlin and Kay's second conjecture was that color lexicons evolve from simple to complex, along highly constrained paths, starting from two BCTs corresponding to warm-or-light and dark-or-cool categories in the simplest lexicons and ending with the 11 BCTs of languages like English.

## The first conjecture: The basic color terms and their universality

Berlin and Kay proposed that most world languages include a set of BCTs in their lexicons. According to

Citation: Lindsey, D. T., & Brown, A. M. (2014). The color lexicon of American English. *Journal of Vision*, 14(2):17, 1–25, <http://www.journalofvision.org/content/14/2/17>, doi:10.1167/14.2.17.

their definition, the BCTs are monolexemic (single, noncompound words that lack modifying prefixes or suffixes) and are used principally in reference to the colors of things, without constraint as to what thing is being described. Moreover, BCTs are present in the idiolects of all informants speaking a given language, are used in a consistent way across all informants, and can be used to partition color space exhaustively. By these and other criteria, Berlin and Kay proposed 11 English BCTs: black, white, red, yellow, green, blue, brown, orange, pink, purple, and gray. In contrast to the BCTs, most languages, including English, have additional color terms that fail to meet one or more of these criteria: Either not everybody uses them (e.g., *chartreuse* in English), they are not monolexemic (e.g., *light blue* in English) or they are restricted as to what they can name (e.g., *blond* in English).

BCTs name basic color categories. While the terms themselves are specific to a language (the same samples might be called *red* in English, *rouge* in French, *akai* in Japanese, and so forth), Berlin and Kay's first conjecture was that the color categories these terms refer to are universal across languages. While not every language has every color category named within its lexicon, Berlin and Kay proposed that "a total universal inventory of exactly eleven basic color categories exists from which the eleven or fewer basic color terms of any language are always drawn" (1969, p. 2).

In the years since Berlin and Kay's work, an enormous amount of research has been done to determine whether their first conjecture is correct. Some investigators have addressed the issue of whether the inventory of terms listed, and only those color terms, fulfill Berlin and Kay's definition of BCTs in every known language. This literature as a whole suggests that at least some BCTs exist in every language that has been examined, but that there might be more than 11 of them in some cases. Boynton and Olson (1987, 1990; for a review, see Boynton, 1997, pp. 144–145) used performance-based measures of color naming to evaluate the special status of Berlin and Kay's English BCTs. When Boynton and Olson's American English-speaking subjects were allowed to use any monolexemic terms to name colors in the Optical Society of America (OSA) Uniform Color Space, they used BCTs with significantly greater speed, consensus, and consistency than nonbasic terms, much as Berlin and Kay predicted. Boynton and Olson also noted that a 12th term—*peach*—might, over time, assume BCT status in terms of naming speed, consensus, and consistency. Sturges and Whitfield (1995) found similar results using Munsell color samples and British English-speaking subjects, except that they suggest that the 12th term might be *cream* (Sturges & Whitfield, 1997). When Uchikawa and Boynton (1987) applied the methods of

Boynton and Olson to Japanese color terms, the results were generally similar. Particularly, Uchikawa and Boynton also found that similar terms—*hada* (skin), meaning tan, and *mizu* (water), meaning light blue—may be making their way into the Japanese basic color lexicon. Similarly, several investigators have discussed terms for *light blue*, which might be a 12th BCT in several other languages (Al-Rasheed, Al-Sharif, Thabit, Al-Mohimeed, & Davies, 2011; Borg, 2007; Friedl, 1979; Ozgen & Davies, 1998; Thierry, Athanasopoulos, Wiggert, Dering, & Kuipers, 2009; Winawer et al., 2007).

The issue of consensus, which was central to the definition of the BCTs, has turned out to be unexpectedly complex. Boynton and Olson (1987) discovered clear individual differences among their observers, and Uchikawa and Boynton (1987) found that there were no 100% consensus colors among the 430 OSA samples corresponding to the Japanese colors *akai* (red), *kuroi* (black), *kiroi* (yellow), and *aoi* (blue). Similarly, Sturges and Whitfield (1995) reported no 100% consensus samples for yellow, pink, orange, and white. Furthermore, the issue of consensus is complicated by the existence of synonyms for many colors: Different words might be used by different informants speaking the same language to name the same or highly similar color categories (*violet* and *purple* might be synonyms in English, *hairoi* and *guree* (gray) might be synonyms in Japanese).

Moreover, Lindsey and Brown (2006, 2009) have shown that, strictly speaking, high consensus may be the exception rather than the rule among the color lexicons in world languages. They examined the World Color Survey (WCS) data set (Kay, Berlin, Maffi, Merrifield, & Cook, 2010), a large database of color naming by 2,616 informants, each speaking one of 110 unwritten languages and living a traditional lifestyle far from daily influences of modern technology. Each WCS informant was tested with a standard set of 330 Munsell color samples of varying hue, value, and chroma (shown in Figure 1a), one at a time, in a fixed pseudorandom order, and provided a color name for each. Lindsey and Brown (2006) used cluster analysis to extract a glossary of universal terms used by WCS informants. A second cluster analysis (Lindsey & Brown, 2009) on the glossed color-naming systems of WCS informants revealed that the color vocabularies of WCS informants clustered into four distinct vocabulary types ("motifs"), where each motif had its own characteristic set of color terms. Crucially, multiple motifs occurred side by side within most WCS languages. This meant, for example, that some speakers of a language might use only color terms glossed as *black*, *white*, and *red*, while others might use five color terms, and still others might use 10 color terms. The lack of consensus revealed by Lindsey and Brown's

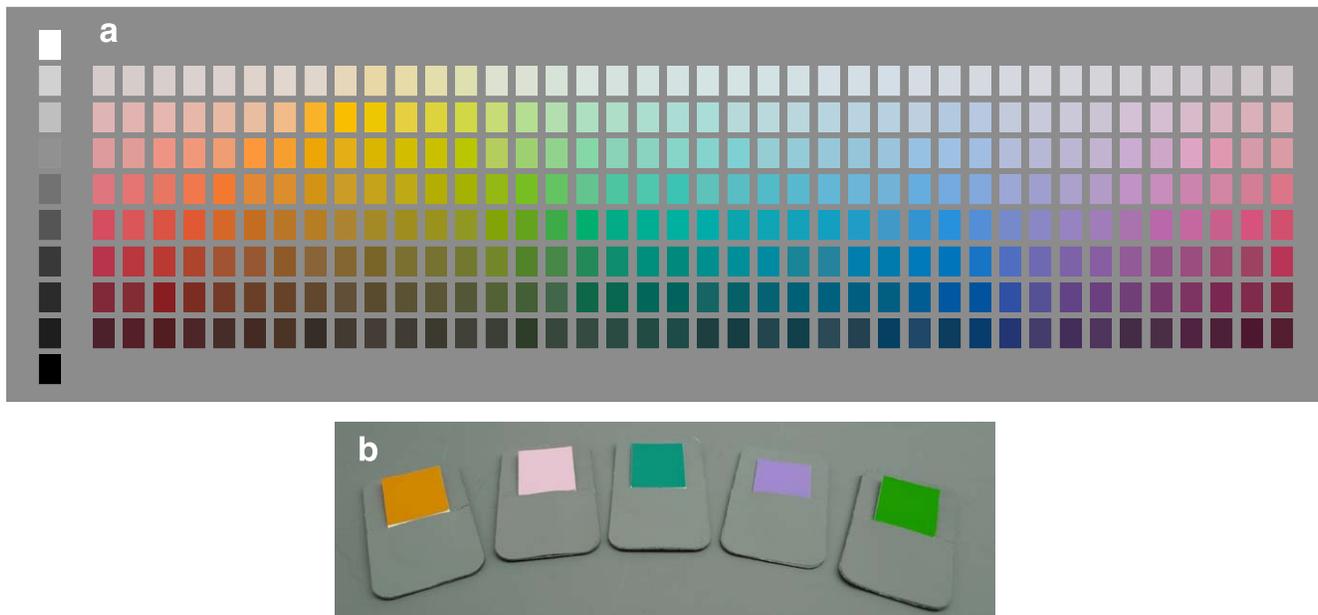


Figure 1. (a) Color chart approximating the samples used in this study, arranged in the order of their Munsell hues (horizontal direction of the chart) and values (vertical direction). (b) Sample holders, shown on Color-aid N4.5 background.

2009 analysis was not merely quibbling about where the boundaries are located in the stimulus set. Rather, it indicates a profound failure of consensus among the speakers of most WCS languages.

The work of Lindsey and Brown (2009) revealed a new kind of universality in addition to the one proposed by Berlin and Kay (1969). Whereas the WCS languages differed from one another in how many individuals used each of the four motifs, the motifs themselves occurred worldwide, even though the informants who used each of the motifs spoke languages with no known historical linguistic ties. This suggested that analysis of color naming must be conducted at the level of each informant's idiolect, rather than at the level of the language shared among a community of informants. Furthermore, the results revealed the usefulness of cluster analysis as an objective means of comparing color naming across languages, thus avoiding many of the pitfalls associated with glossing color terms by traditional lexicographic techniques.

### The second conjecture: The evolution of BCTs

Berlin and Kay's second conjecture was that languages evolve over time by adding new color categories (see also similar concepts proposed by Gladstone, 1858; Rivers, 1901; Hugo Magnus, translated in Saunders & Marth, 2007; and Schontag & Schafer-Priess, 2007). According to the second conjecture, new color terms are continually being added

to languages that have fewer color terms by "successive differentiation of previously existing color categories" into smaller, more accurately named subcategories (Kay & McDaniel, 1978, p. 640). This process follows a series of stages, in a fairly constrained evolutionary trajectory. According to the second conjecture, this continues until the lexicon reaches a stage equivalent to the 11 BCTs of English and other languages spoken in industrialized societies. Thus, Berlin and Kay's second conjecture is that color lexicons evolve, that they follow a prescribed trajectory, and that color terms are added by "partition" of existing named color categories.

In a sense, the strict ordering proposed by Berlin and Kay resembles a theory of biological development, in which maturation of the organism occurs in stages following a single prescribed trajectory, with minor differences from individual to individual. More recent work by Kay et al. (2010) has relaxed and generalized the evolutionary hypothesis considerably, to allow for some languages that do not fit neatly into one of Berlin and Kay's original stages and to suggest a much less constrained, more diverse range of evolutionary pathways. The more diverse range of trajectories proposed by Kay et al., 2010 resembles ontological evolution more closely, where a population of organisms can evolve in any of a number of directions, subject to the Darwinian principles of natural selection.

Berlin and Kay's second conjecture poses two important questions pertaining to languages like American English, which are spoken in technologically advanced cultures. First, is the current state of modern

color naming in these languages a proper end state of color-term evolution, as Berlin and Kay proposed? There is some evidence that it is not. First, several modern languages with more than 11 basic color terms have been identified, which otherwise satisfy Berlin and Kay's criteria. These include Russian (Davies & Corbett, 1994; Winawer, 2007), Greek (Thierry et al., 2009), and Turkish (Ozgen & Davies, 1998). Second, there is some evidence that other English-like color lexicons may be continuing to evolve. For example, Zollinger (1984) proposed that *turquoise* may be a nascent color category in German, and Boynton and Olson (1987) proposed that even English itself might be currently evolving, adding *peach* as a possible new color category.

The second question arises if we grant the likelihood of the continuing evolution of the English color vocabularies: How are the new categories formed? Is the process constrained by the partition principle of Kay et al., or by some other process? An alternative process has been advanced by Levinson (2000) and Lyons (1995), who have challenged both of Berlin and Kay's conjectures. Here, we focus on the second conjecture and Levinson's idea that in the earliest stages of color-term evolution, color vocabularies do not exhaustively name all colors. Rather, according to Levinson, each ancient color term referred to a restricted range of colors that were identified with a particular item in the environment, for example a certain animal or plant, or a substance such as blood or bile. Over time, the original terms generalized to the colors of the substances to which they originally referred. However, great gaps remained where the colors were either unnamed or else were named with great difficulty. According to Levinson, additional color terms came into use as the need arose to name colors in the gaps, colors that previously had no names. Thus, according to Levinson, "color terms [emerge] out of noncoloric expressions" (2000, p. 8); this view of color-term evolution has come to be known as the "emergence hypothesis."

## This project

In the present study, we examined Berlin and Kay's two conjectures in light of a new color-naming data set that we obtained from American English-speaking informants. To facilitate comparisons between American English and the 110 languages represented in the World Color Survey, we used the set of Munsell color samples used in the WCS. We used a "free-naming" protocol, in which informants used whatever single color term they wished, subject to a few simple rules, and we added a "constrained-naming" phase to the data-collection protocol, in which only the 11 BCTs of

Berlin and Kay were allowed. This constrained-naming phase of the protocol was used to establish each informant's BCT category boundaries. Then, the deployment of nonbasic color names—within versus between BCT categories—in the free-naming phase could be gauged in relation to these boundaries.

The data analysis in the present project followed the two-stage cluster-analysis methodology of Lindsey and Brown (2006, 2009). In the first stage, each color term used by each informant was encoded by a separate binary feature vector, which represented the subset of color samples associated with each color word. These feature vectors were then partitioned into distinct clusters, which represented a glossary of distinct color categories identified in the data set irrespective of the actual words associated with the clustered feature vectors. This step avoided the potential pitfalls of synonymy in the data analysis. In the second stage, the clustered feature vectors were used to reconstruct a representation of the glossed color-naming system for each American English-speaking informant, and a second cluster analysis of the data set was performed at the level of the informants. This step allowed us to determine whether distinct subpopulations of American English-speaking informants express different color-naming motifs. This two-stage cluster analysis permitted an analysis of the American English color lexicon that was more nuanced and powerful than one based on simple tabulation of subject color-naming responses.

This analysis was designed to address the issues outlined previously. Under the first conjecture, are there color terms in American English that fulfill Berlin and Kay's definition of the BCTs? And are the most common American English color terms equivalent to the universal color terms that have previously been identified for the WCS, or do they differ from those universal terms in important ways? Under the second conjecture, is American English in the process of evolving to higher numbers of BCTs, as suggested by Boynton and Olson? And if new color terms are evolving in American English, do they appear by partitioning existing categories into smaller ones, as Kay et al. proposed? Or is Levinson closer to the mark, with these new terms appearing *de novo*, popping up in places where no high-consensus color term exists and informants find the colors hard to name? Finally, we used the data set to examine the relationship between informant gender and color naming. Several studies have proposed that females have larger color vocabularies than males, possibly for genetic reasons, and women might have a finer appreciation of the differences between colors and their identities because of their roles in modern and traditional cultures.

This data set also allowed us to examine prospectively the generality and replicability of the motifs of Lindsey and Brown (2009). Their conclusion that multiple motifs exist side by side within the lexicons of most world languages was based on an analysis of the WCS, a data set that had already been collected. This unexpected result suggested that the color lexicons of many world languages are currently undergoing linguistic change, albeit over more trajectories than the simple path originally proposed by Berlin and Kay. Therefore, it was important to determine whether multiple motifs occur in a data set that was collected to reveal them if they exist. On one hand, American English is a written language spoken by a large number of people, which suggests that it might not still be evolving in a basic, common lexicon such as the names of colors. On the other hand, American culture is highly industrialized, which suggests that there might be a need for more color terms as more and more artifacts in American culture differ from one another only in their color.

## Methods

### Informants

Fifty-one native American English-speaking informants (24 men and 27 women; ages 19–58 years) participated in the study. All were born and raised in the United States, none had spoken any language other than English at home before the age of 12 years, and all resided in the Columbus, Ohio, metropolitan area at the time of testing. None of the informants were aware of the hypotheses being tested in this study. All subjects reported that they were free of visual pathology except for refractive error, and all were color normal, as assessed using the D-15 screening test. The informants were tested following a protocol previously approved by the Ohio State University's Institutional Review Board, and all gave informed consent prior to participating in this study.

Data on three additional male informants, one who spoke British English and two who were color-vision specialists and well aware of the hypotheses being tested in this study, have been eliminated from the data set reported here. None of the results or conclusions from this project are materially affected by the exclusion of these informants.

### Color samples

The 330 color samples were taken from the *Munsell Book of Color Glossy Edition* and corresponded closely to those used in the World Color Survey color chart

(Kay et al., 2010). Ten samples were achromatic, with values from 1.5/ to 9.5/ in the notation of the Munsell color-order system. The remaining samples were the 40 equally spaced Munsell hues (2.5 R to 10 RP, in hue steps of 2.5) sampled at each of eight values from 2/ to 9/ (hence, 320 hue/value combinations). The chromatic samples generally had high chroma, except for some hues at the lowest and highest values, where the *Munsell Book of Color* does not contain high-chroma samples. Each sample was placed in a 51-mm × 28-mm holder, which was covered in Color-aid N4.5 paper exposing a 20-mm × 20-mm (2.3° × 2.4° at an approximate viewing distance of 500 mm) portion of its Munsell sample.

### Apparatus

The color samples were presented in a custom-made light box, with white walls, a floor covered with Color-aid N 4.5 paper, and a 42-cm × 142-cm opening through which subjects viewed the illuminated color samples. Illumination was provided by a bank of four full-spectrum fluorescent lights (F40T12 Spectralite; CRI 90) suspended from the top of the light box. Color calibrations were performed using a PhotoResearch PR-670 spectrophotometer at regular intervals throughout the duration of the study. Illumination varied between 1970 and 2216 lux during the course of the study and had a correlated color temperature (CCT) of between 5200 and 5400 K during the same time period. This CCT is near that of direct sunlight.

### Procedure

At the beginning of each experimental session, the informant was briefed on the nature of the task. After the informant provided informed consent for participation in accordance with the Declaration of Helsinki and under the approval of the Ohio State University Institutional Review Board, we obtained the informant's age and sex, the languages he or she spoke and at what age he or she learned them, where he or she was born and lived as a child, and where he or she lived presently. We also administered the D-15 panel to screen for color-vision deficiency.

The color-naming part of the experiment consisted of two phases: free naming and constrained naming. The *free-naming* phase of the experiment was always completed first, and informants were not apprised of the second, *constrained-naming*, phase until after the first phase was completed. All but one informant completed both color-naming phases in one 1.5-hr experimental session. The remaining informant required two 1.5-hr sessions to complete the two color-

naming phases of this study. Subjects were paid with a \$10 gift card for their participation.

In the free-naming phase of the experiment, the experimenter presented each of the 330 Munsell samples in the light box in a fixed, pseudorandom order. The subject named each color sample in turn. The instructions were to name the colors of the samples, based on the following three criteria:

1. The color name must be a single word. (Phrases like *light blue* and *dark green*, and phrases with intrinsic modifiers like *yellowish* are not acceptable.)
2. The word must be a general color name, applicable to anything of that color. (*Blond*, for example, is not such a word, as it is used to name the color of hair, furniture, or beer, but not, for example, a car or a potato.)
3. The word must be the one that you would normally use to name the color of something in your everyday life. (We are not looking for a unique name for each color. We are not testing for how many different color names you know or can dream up, or how many subtle distinctions in color you can name. We want just to know how you naturally name the colors, when you can use only a one-word name.)

Once we had instructed the informant and answered any questions, the informant was allowed to use whatever color terms he or she chose: Informants complied with criterion #1 without exception, and we did not interrupt to object to terms that might have violated criteria #2 or #3. Color terms that were not among the 11 BCTs of Berlin and Kay were flagged on the data sheet. The free-naming data set was the full set of monolexemic color terms provided by this sample of informants.

The instructions in the free-naming phase of data collection differed somewhat from those used in the WCS protocol. In that study, field-workers were encouraged to elicit short, single-word BCTs from their informants, in their native language. However, there was probably considerable variation in how these instructions were actually followed by field-workers (see Cook, Kay, & Regier, 2005). Thus, we believe our color-naming protocol adhered reasonably well to that of the WCS as actually implemented in the field. Moreover, by using a relatively unconstrained color-naming protocol in the first phase of data collection, we could compare our results to those of prior studies of English monolexemic color naming by Boynton and Olson (1987, 1990) and by Sturges and Whitfield (1995, 1997).

In the constrained-naming phase of the experiment, we gave the informant a printed list of the 11 BCTs. We then presented, for a second time and in reverse order, the samples corresponding to the color terms that were flagged in the free-naming phase of the experiment, and

required the informant to provide a BCT from the list. The colors provided in the constrained-naming phase were added to the data set obtained in the free-naming phase to create a second complete constrained-naming data set.

## Results

### Color terms elicited under free-naming instructions

Every informant succeeded in naming every sample with a monolexemic color term in the free-naming phase of the study. The 51 native American English-speaking informants used a total of 122 color terms to name the 330 Munsell samples (Tables 1 and 2). Figure 2a and Table 3a present the most commonly used basic and nonbasic color terms in this study. The mean number of free-naming color terms per informant was 21.9 ( $SD = 7.6$ ) and the minimum number of color terms was 12 (Figure 3a).

Among the nonbasic terms, *peach* was used by the largest number of informants (40 out of 51 informants). *Teal*, which was used by 32 of 51 informants, was the only other nonbasic term used by over half of informants. Fifty-nine of the 122 color terms (48.4%) were used by only one informant each, and of those, 25 were used only a single male informant. The inventory of frequently used nonbasic color terms in Figure 2a and Table 3a is qualitatively similar to that reported by Boynton and Olson (1987, 1990) and by Sturges and Whitfield (1995). However, none of those previous studies mentions *teal* in its nonbasic color inventory; instead, they all report the use of *turquoise*, which we found to be synonymous with *teal* (we deal with color-term synonymy later). None of Sturges and Whitfield's informants used *lavender*, though they did report the term *lilac*, which we found to be synonymous with *lavender*. Finally, Boynton and Olson (1990) reported that both *peach* and *tan* were used by all nine of their informants, compared to 62.7% and 45.1%, respectively, of informants in this study. Boynton and Olson (1987, 1990) speculated that *peach* might be an emerging 12th English BCT. We will return to this point in the Discussion section of this article.

While every informant used all the BCTs of Berlin and Kay, the BCTs differed considerably in their frequency of usage (Table 3a). On average, 24.0% of the sample presentations (4,035 of 16,830 total responses) elicited *green* as a color term, followed by *blue* (16.9%). *Red*, with a usage rate of 3.3%, was the least elicited of the chromatic BCTs. Of the neutrals, *black* was elicited by 1.9% of the 330 color samples across the 51 informants, *white* was elicited by 4.7%, and *gray* was

Study	Number of informants	Number of color samples	Type of samples	Number of terms	Terms/informant				
					Mean	SD	Median	I.q.r.*	Mode
Present study	51	330	Munsell (World Color Survey)	122	21.88	7.58	20	7.5	18
Boynton and Olson (1990)	9	424	OSA	82	32.89	14.84	31	29.5	N/a**
Sturges and Whitfield (1995)	20	446	Munsell	N/a***	N/a***	N/a***	N/a***	N/a***	N/a***

Table 1. Basic data from three studies. Notes: \*Interquartile range. \*\*No two subjects used the same number of terms. \*\*\*Not reported.

elicited by 2.2%. The nonbasic term used to name the most samples was *teal* (2.0% of samples), followed by *peach* (1.5%) and *lavender* (1.3%). The only term for *light blue* was *sky*, which was used by only four subjects, to name 0.21% of samples.

Many investigators have found that the frequency of word use conforms to a power law—that is, the logarithms of the frequency with which words occur fall on a line when graphed as a function of the logarithm of their rank order. This power-law relation is sometimes called Zipf’s law (see Mitzenmacher, 2003, for a review). Contrary to that general result, when we graphed the number of informants using each term (the term’s “popularity”) as a function of the sorted rank order of that term’s popularity (Figure 4, lower data set), the data were broken quite sharply into three regimes. First, there was a ceiling effect, as the BCTs were all used by all 51 informants and are therefore fitted perfectly by a constant function. For the next 17 most popular terms, the power law had an exponent of  $-1.2$ , whereas the power law for the less popular terms was  $-3.32$  (gray circles in Figure 4). The slopes of these functions depended somewhat on how ties were treated in the rank ordering (here, tied frequencies have consecutive ranks) and how the BCTs were shown on the graph (here, included as 11 tied frequencies). However, no matter how we treated the ties, there was always a break in the function after the 28th term (11 BCTs plus 17 additional popular terms; numbers listed in Table 2; colors listed in Table 3a). Double-power-law behavior is common in language corpora, where two exponents often “divide words in two different sets: a kernel lexicon formed by about  $N$  versatile words and an unlimited lexicon for specific communication” (Ferrer i Cancho & Solé, 2001, p. 170).

Criterion	All	Chromatic	Nonbasic
Used by > 0 informants	122	122	111
Used by > 1 informant	63	63	52
Used by > 2 informants	51	51	40
Used by > 3 informants	43	43	32
Basic color terms (Berlin & Kay, 1969)	11	8	0
Most common color terms*	28	25	17
Glossed color categories**	20	17	9

Table 2. Free-naming color terms. Notes: \*White disks in Figure 4. \*\*See Figures 2b and 3b and Table 3.

### Color terms elicited under constrained-naming instructions

Under the constrained color-naming instructions, all the informants succeeded in naming all of the samples for which they had provided nonbasic color terms in the free-naming part of the protocol. This was

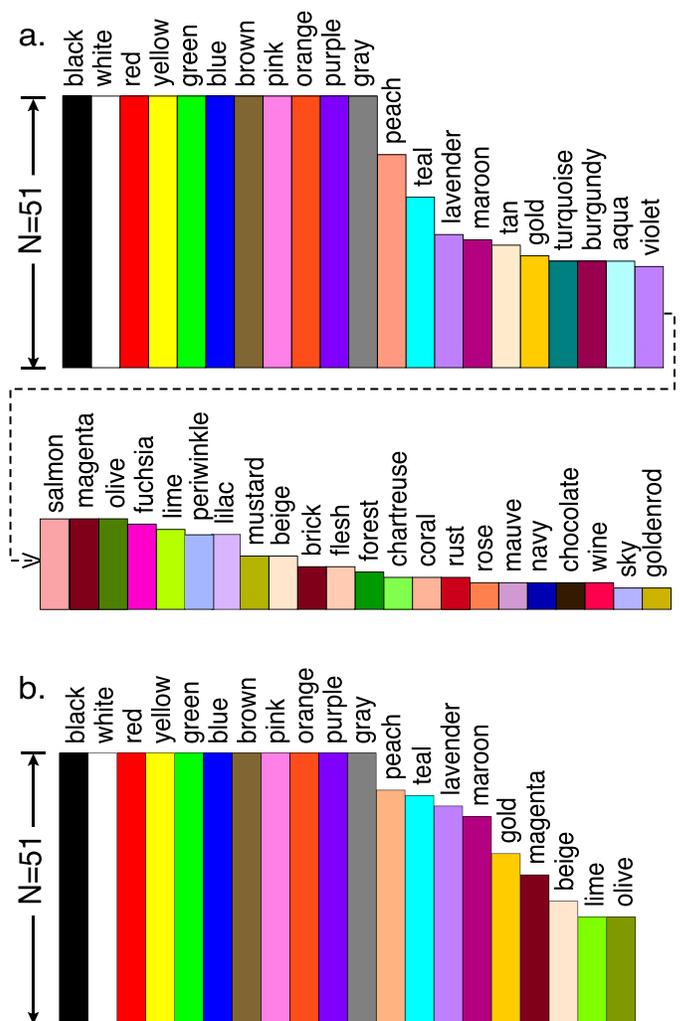


Figure 2. Histograms of color-term usage in the free-naming phase of the experiment. (a) The number of informants using each of the 43 color terms used by four or more informants. (b) The free-naming data consolidated by cluster analysis into glossed categories.

	Rank by popularity	Number of informants (popularity)	Number of samples (usage)
<b>(a) Unglossed color terms</b>			
GREEN	1	51	4,035
BLUE	2	51	2,843
PURPLE	3	51	1,668
PINK	4	51	1,302
BROWN	5	51	930
WHITE	6	51	788
ORANGE	7	51	716
YELLOW	8	51	628
RED	9	51	558
GRAY	10	51	375
BLACK	11	51	315
PEACH	12	40	247
TEAL	13	32	338
LAVENDER	14	25	215
MAROON	15	24	109
TAN	16	23	58
GOLD	17	21	100
AQUA	18	20	136
TURQUOISE	19	20	131
BURGUNDY	20	20	72
VIOLET	21	19	163
OLIVE	22	17	112
MAGENTA	23	17	84
SALMON	24	17	40
FUCHSIA	25	16	66
LIME	26	15	40
LILAC	27	14	81
PERIWINKLE	28	14	57
SKY	42	4	35
<b>(b) Glossed color categories</b>			
GREEN	1	51	4,154
BLUE	2	51	2,997
PURPLE	3	51	1,731
PINK	4	51	1,366
BROWN	5	51	962
WHITE	6	51	788
ORANGE	7	51	752
YELLOW	8	51	655
RED	9	51	591
GRAY	10	51	375
BLACK	11	51	315
PEACH	12	44	312
TEAL	13	43	589
LAVENDER	14	41	447
MAROON	15	39	232
GOLD	16	32	174
MAGENTA	17	28	119
BEIGE	18	23	86
OLIVE	19	20	127
LIME	20	20	58

Table 3. Usage and popularity of color terms and categories. (a) Unglossed color terms ranked 1–28 in the order of their popularity (white circles in Figure 4), plus sky (see Discussion for details). (b) Glossed color categories (triangles in Figure 4).

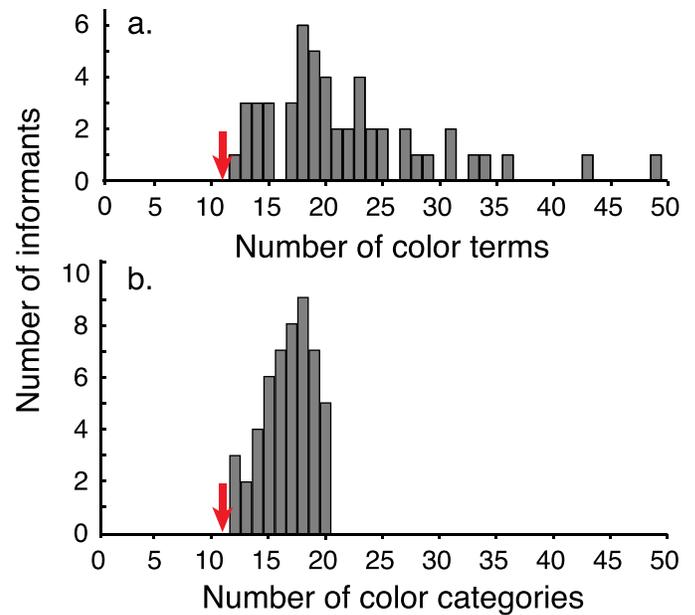


Figure 3. (a) The distribution of the number of informants using each total number of color terms in the free-naming phase of the experiment. (b) The distribution of the free-naming data after synonymous color terms were consolidated into color-category clusters. The red arrows indicate 11, the number of BCTs according to Berlin and Kay (1969). Even after consolidation, no informant used only 11 color terms, and the modal number of glossed color terms was 18.

accomplished with small amounts of grumbling from some informants, who found some of the peach-colored samples particularly challenging.

All informants used all of the BCTs to name at least some of the color samples, so the median/average number of basic color terms was 11, and the interquartile range/standard deviation was nil. On average, 27.9% of samples (4,693 of 16,830 total responses by 51 subjects) elicited *green* as a color term, followed by *blue* (19.3%), with *red*, once again, being the least used chromatic term (4.2%). *Black* was elicited by 1.9% of total responses, *white* was elicited by 4.9%, and *gray* was elicited by 2.3% of trials.

### Consensus in color-term usage

In their 1969 monograph, Berlin and Kay wrote that the BCTs in any language, including English, are “psychologically salient.” By this, they meant that the BCTs are used with high consensus and consistency by all competent speakers of the language. What were the salient colors in this data set? Figure 5 shows consensus maps for the free-naming (left) and constrained-naming (right) phases of this study. Each of the 330 small rectangles within each map in Figure 5 corresponds to a color sample (its true color is illustrated in the

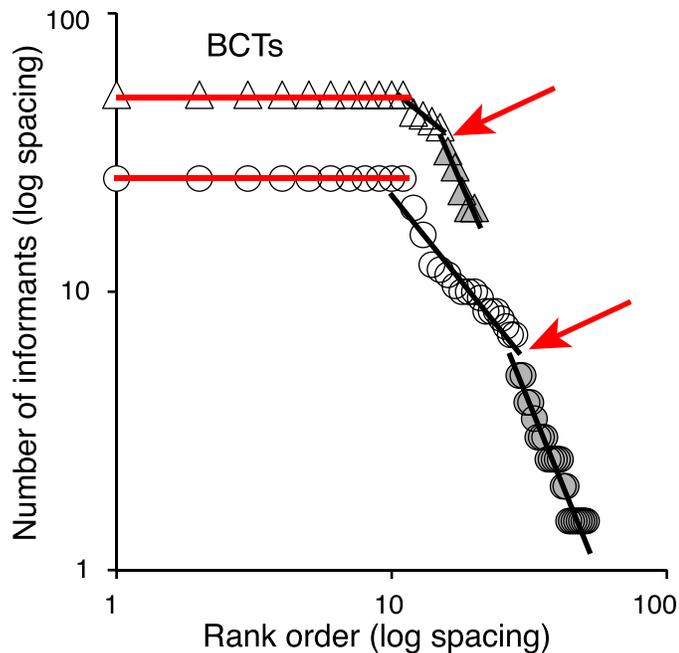


Figure 4. The number of informants using each of 43 unglossed color terms used by four or more informants (circles), and 20 glossed color terms (triangles), plotted against their sorted rank order. The first 11 terms, used by all 51 informants, are the BCTs (horizontal red lines). Both graphs show reliable changes in slope (arrows), with rare color terms being used by very few informants (gray symbols). Circles are displaced downward for clarity.

corresponding location in Figure 1a), and its false-color hue encodes its most frequent color term. In the case of *black*, the false-color code was *orange* (the actual *orange* category is in a lighter shade of orange in the figure), and for *gray* it was *chartreuse*. The BCTs constitute the most frequent names for 99% (326/330) of the color samples named by the informants in the free-naming phase of this study. Only four color samples, which were in the peach area of the color chart, received nonbasic color terms most of the time. The maximum consensus for *peach* was 0.61 of informants (see Boynton & Olson, 1987; Jameson & Alvarado, 2003, for a similar result). Of course, the BCTs were the modal color terms for all the samples in the constrained color-naming data set.

The false-color light-to-dark values of the samples shown in Figure 5a and d encode the consensus—that is, the fraction of informants who used the modal name, normalized to the overall maximum value for visibility in the figure. The consensus in the constrained color-naming data set (Figure 5d) was higher within color categories and lower near the color-category boundaries, indicating variation across subjects in the locations of the boundaries between the BCTs. The consensus was generally lower overall in the free-naming task (average consensus = 0.74; Figure 5a) than in the constrained-naming task (average consensus = 0.85), because informants used nonbasic terms in addition to the BCTs in the free-naming task.

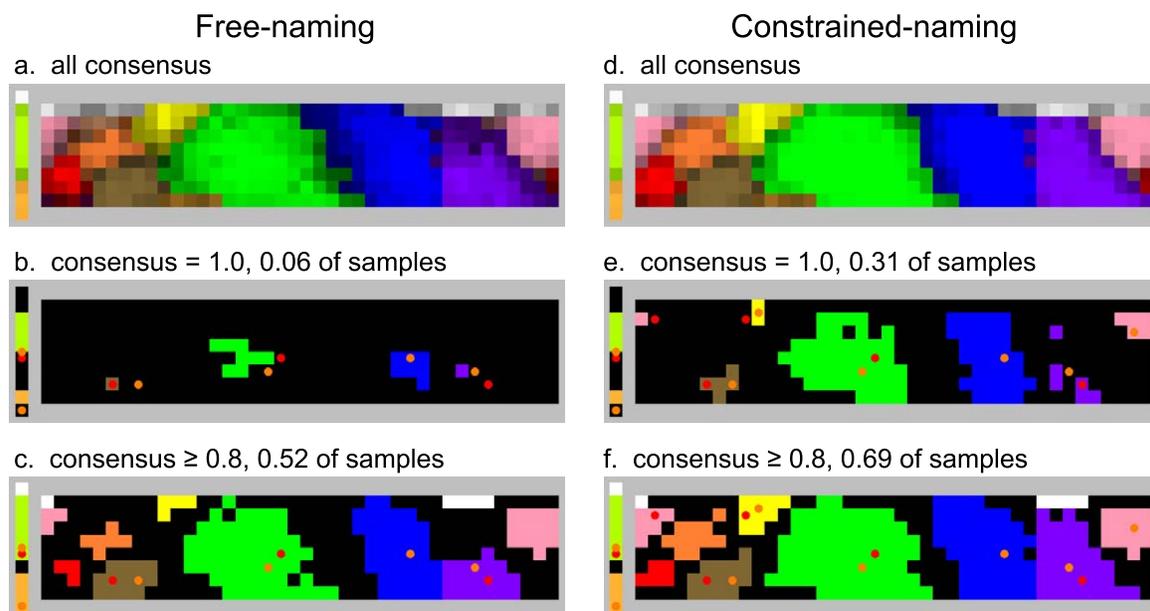


Figure 5. Diagrams of color-naming consensus in the free-naming (a–c) and constrained-naming (d–f) tasks. See text for details of color coding. (a, d) Consensus maps of the usage of the BCTs of Berlin and Kay (1969). Second and third rows indicate samples for which the consensus reached or exceeded two threshold criteria: 1.0 (b, e) and 0.8 (c, f). Red dots: focal colors from Berlin and Kay (1969); orange dots: focal colors from Sturges and Whitfield (1995).

In Figure 5b, c, e, and f, the data from Figure 5a and d were examined by applying two threshold levels of consensus: 1.00 (Figure 5b, e) and 0.80 (Figure 5c, f; 0.80 was one of the criterion levels used by Kay et al., 2010, in their description of the World Color Survey data set]. In those diagrams, the chromatic color categories named at or above the critical value of consensus appear as islands of at-or-above-threshold agreement, separated by black boundary regions of below-threshold agreement. Agreement among subjects was perfect (consensus = 1.0) for 6.3% of samples in the free-naming data set and for 31% of samples in the constrained-naming data set. Fifty-two percent of the samples in the free-naming data set and 69% of the samples in the constrained-naming data set reached or exceeded a consensus criterion of 0.80. In the free-naming data set, none of the samples were called *white*, *red*, *yellow*, *pink*, or *orange* by all informants, and even in the constrained-naming data set, no samples were called *white* or *red* by all informants. However, at the 0.80 consensus threshold, all the BCTs were represented in both the free-naming and constrained-naming data sets. The dots in Figure 5 are the “focal colors” of Berlin and Kay (1969) and Sturges and Whitfield (1995), that is, the colors that were chosen by their informants to be the best examples of each of the color categories. The focal colors correspond closely to the regions of high color-naming agreement, which indicates that color samples were named with high consensus if they were, on average, particularly good examples of named color categories. Conversely, as colors deviated more and more in appearance from the focal colors, subjects were more variable in their responses and consensus agreement declined.

The features of the 0.80 threshold consensus map for the constrained-naming task generally agree with the results of prior studies of English color naming by other investigators (Boynton & Olson, 1987, 1990; Sturges & Whitfield, 1995). Green and blue were the only named categories that extended vertically throughout the lightness range. All the other basic color categories were confined to restricted ranges of lightness: Pinks and yellows were light compared to the neutral background against which the colors were viewed, and reds and browns appeared among the lower lightness warm colors; orange was at an intermediate lightness value.

### Color-term centroids

To examine the individual data, we calculated the color-term centroids for each term for each informant from the free-naming phase of the study (Figure 6). Each centroid (shown by the colored dots)

was defined as the average coordinates of the samples that were named with the corresponding term by one subject, specified within the 2-D Cartesian coordinate frame of the color chart shown in Figure 1. The faint colors in the backgrounds of Figure 6a through d are the false colors from Figure 3f, corresponding to the modal BCTs used at consensus  $\geq 0.80$ . Here, we provide this map as a guide for examining the color-term centroids. The black dots in Figure 6a are the averages of the centroids across all 51 subjects. For comparison, the white dots are the average centroids obtained from 20 University of Teesside (U.K.) undergraduates by Sturges and Whitfield (1995), whose 446 Munsell color samples spanned a greater range of chromas than were used in the present study. The centroids from the two studies agree fairly well.

A striking feature of the individual data is the informant-to-informant variation in the usage of the nonbasic color terms. Informants often used different color names to label similar regions of color space. For example, in the central region of the color chart (Figure 6b), there were seven different nonbasic color terms: *teal*, *turquoise*, *aquamarine*, *aqua*, *jade*, *ocean*, and *seafoam*. These terms reliably denoted colors that fall near the boundary between the blue and green BCT categories, and the distributions of the centroids for these seven terms were broad and showed considerable overlap. However, the centroids were not quite identical, suggesting that informants might differ slightly in the meanings they associate with these nearly synonymous color terms. The centroids for *maroon* and *burgundy* showed almost perfect overlap, and the centroids for *lavender* and *lilac* also overlapped greatly within the upper lightness range of samples called *purple* on the constrained-naming task. The centroids for *violet*, like those for *teal*, covered a large range of lightnesses, suggesting that *violet* was generally not synonymous with *lavender* and *lilac*.

In contrast, some color terms were used to name quite different colors by different individual informants. For example, the centroids for *tan* (Figure 6b) appeared in two disjoint areas of the color diagram: One area overlapped with the centroids for *peach* and *beige*, and the other was close to the centroids for *olive*. Similarly, informants used *chartreuse* to mean either greenish-yellow or desaturated green (synonymous with *lime*). *Puce* is also interesting, as all three of the informants who used this term applied it to yellowish-green-colored samples. Apparently informants using *puce* did not know that it refers to a dark, highly saturated purplish red or purplish-brown, and they assigned the color term instead to the color of vomit (purple dots in Figure 6b).

## The nonbasic color terms: Partition or boundary colors?

These data sets allowed us to examine the second conjecture and the two hypotheses about the origins of new color categories: the partition hypothesis of Kay et al. and the emergence hypothesis of Levinson. We examined the locations of the centroids of the color-naming patterns in the free-naming data set relative to the boundaries of the BCTs obtained from the constrained-naming data. If the partition hypothesis is correct, then each nonbasic color term will tend to be located within one of the BCT categories, with its centroid at some distance from the nearest BCT boundary. In contrast, if the emergence hypothesis is correct, and new terms intrude into the areas between named categories, then the centroids for the nonbasic color terms will be located near the nearest BCT boundary, and the average distance to the nearest boundary will be near zero.

For each color-naming pattern for each informant, we calculated the unsigned distance between its centroid and the nearest BCT boundary, expressed as the number of samples between the centroid and the closest boundary (above, below, to the right, or to the left of the centroid). Inasmuch as the centroids were not integers, the separation between the centroids and the boundaries were not integers either. Figure 7 shows the results of this analysis in two ways. In the line graphs, the distance data from the informants who used a given color term were binned into half-chip bins. Each line shows, for a given color term, the number of informants who placed it within 0.5 chips of the nearest boundary, between 0.5 and 1 chip from the nearest boundary, and so forth. Not surprisingly, the centroids of the BCTs were well centered within their respective categories (Figure 7a; see also Figure 6a), so the distances to their nearest boundaries were generally greater than 1 chip. The closest bin (under 0.5-chip separation) was never the most frequent separation between a BCT centroid and its nearest boundary. In contrast, the distance data for 13 of the 17 most frequently used nonbasic chromatic colors were at their maximum within 1 chip of zero, as predicted by the emergence hypothesis. Figure 7b shows the results for four representative nonbasic chromatic color terms. However, the distance data of four of the 17 most common nonbasic terms peaked at a distance of more than 1 chip. These were *lavender*, *lilac*, *olive*, and *lime* (Figure 7c; compare to Figure 6c, arrows).

The bar graph (Figure 7d) shows the average of the unsigned nearest boundary distances for each color term. The average distances for the BCTs and the nonbasic chromatic color terms overlap only slightly: Seven of the 25 most frequently used chromatic color terms (the chromatic color terms shown with the white

disks in Figure 4; see Table 2) fall into the overlap region between 1 and 1.5 chips; the average distances for the other colors were all cleanly divided between the BCTs, which were centered in their BCT categories, and the nonbasic color terms, which appeared at the boundaries of the color terms. The three nonbasic terms that fell in the overlap region were *lavender*, *lilac*, and *lime*. The average distance between the BCTs and their nearest color boundary was 1.78 chips ( $SD = 0.89$ ); for the nonbasic color terms it was 0.65 chip ( $SD = 0.34$ ), a statistically significant difference:  $t(7) = 3.50$ ,  $df = 8$ ,  $p = 0.008$ .

In summary, the nonbasic color terms generally appeared at the boundaries between the BCT categories. This result was broadly in agreement with Levinson's emergence hypothesis. However, the color terms for light purple (*lavender* and *lilac*) and light and dark yellowish-green (*lime* and *olive*) appear to be partition colors in the sense of Kay et al. These results show that both processes can occur, although the intrusion of new colors in between established categories may be more frequent in modern American English. Previous work by Sturges and Whitfield (1997) has suggested qualitatively that British English might also have more intrusion colors than partition colors.

## American English glosses

Inspection of the list of color terms from the free-naming phase of the experiment (Table 3a), and examination of the individual data outlined previously, suggested that many of the terms that subjects used might be synonyms. Perhaps the very large number of color terms shown in Figures 2a and 3a would be much smaller if those synonym groups were to be consolidated into larger color categories, much as Lindsey and Brown (2006) did in their cluster analysis of the WCS. Therefore, we applied a similar  $k$ -means analysis to the present data set. Briefly, we expressed each chromatic color term (i.e., each term not used to name any of the 10 achromatic color samples in the WCS chart) deployed by each of the 51 informants as a 320-element binary feature vector, representing the 320 chromatic color samples in the WCS chart. For a given color term (say, *yellow*) used by a particular informant, each element of the term's feature vector was assigned the value 1 if that informant called the corresponding chromatic sample *yellow* and 0 otherwise. Applying this encoding to all of the chromatic words used by our informants yielded a total of 963 binary feature vectors. We then performed a  $k$ -means cluster analysis, computing a partition of the feature vectors into  $k = 2$  clusters, then  $k = 3$ , then  $k = 4$ , and so forth. The  $k$ -means process works by assigning each feature vector to the "nearest" cluster in feature space. We used a

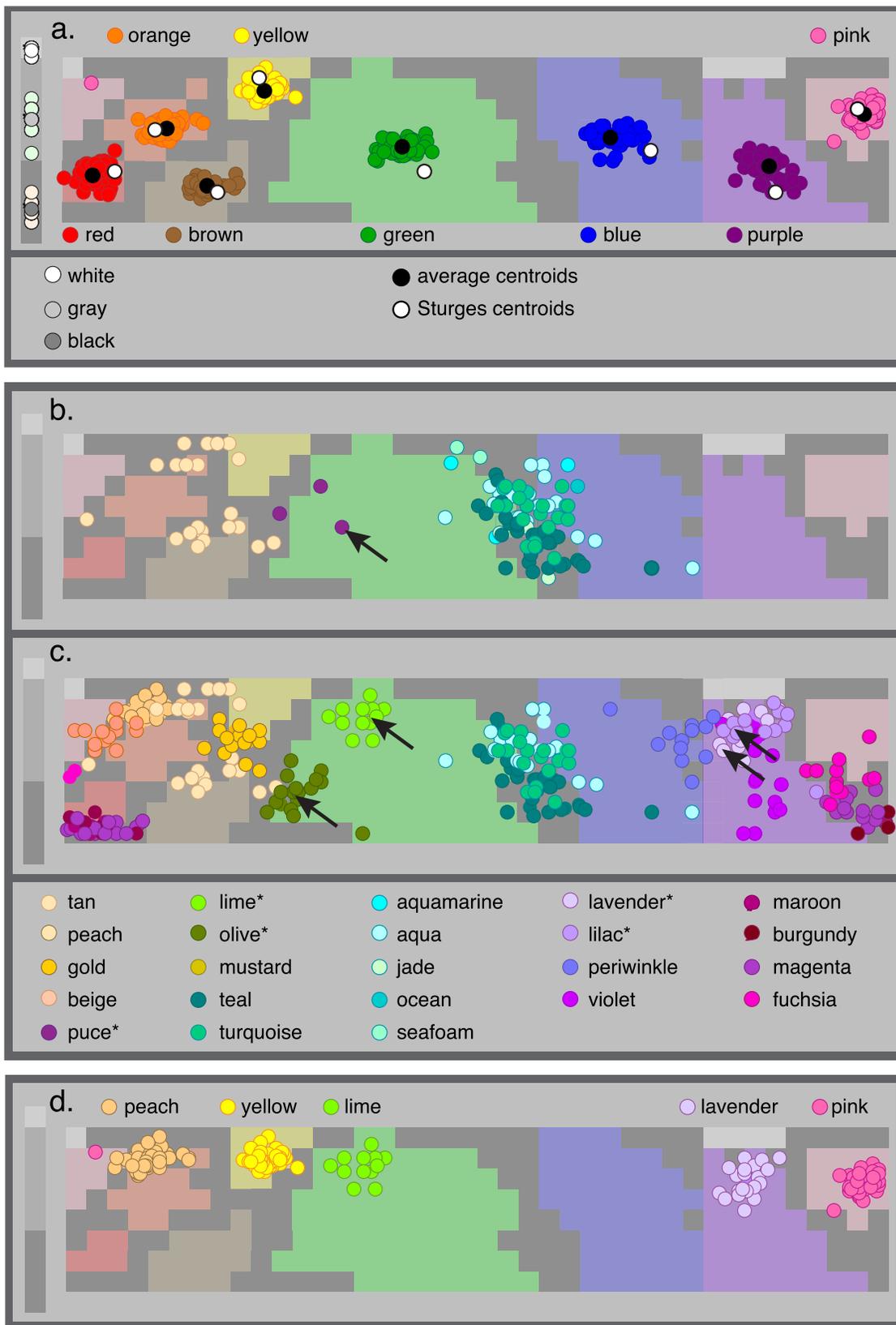


Figure 6. Centroids of named color categories provided by individual informants in the free-naming task. (a) Centroids of the 11 BCTs of Berlin and Kay (1969), with the group average centroids (black disks) and the centroids of Sturges and Whitfield (1995; white disks). Color key for the chromatic color terms is above and below the diagram. (b) Individual differences in usage of nonbasic color terms. The centroids for *tan* are disjointly distributed in the warm-color region of the chart. Informants used several terms—some of which

→

←

were uncommon—in the area between the *green* and *blue* BCTs: *teal*, *turquoise*, *aquamarine*, *aqua*, *jade*, *ocean*, and *seafoam*. The dark purplish centroids within the *green* region (arrow) are for the color *puce*. (c) Centroids of the 17 most common color terms (Figure 4, Tables 2 and 3a). Most nonbasic color terms in the free-naming task fall near the boundaries between the BCTs, where consensus for the BCTs is low. However, *lime* and *olive* and *lavender* and *lilac* (arrows) are generally proper subsets of *green* and *purple*, respectively. Color key for (b) and (c) is below (c), and the asterisks refer to the centroids indicated with arrows. (d) Informants used color terms for the light colors *peach*, *yellow*, *lime*, *lavender*, and *pink*, but not for *light blue*. Color key above the diagram.

distance metric of  $(1 - r)$  to determine nearness, where  $r$  is the Pearson correlation coefficient between a feature vector and a cluster centroid.

The  $k$ -means process is an iterative clustering algorithm, and the initial positions of the centroids in feature space must be assigned. A total of  $k$  feature vectors, chosen at random from the data set, served as the initial centroids at the start of each run. The resulting  $k$ -means partition can be somewhat sensitive to these initial conditions. Therefore, each  $k$ -means partition reported here is based on the best of 100 replications per value of  $k$ , where “best” was defined as the clustering that produced the smallest within-cluster distances among cluster members, summed across all  $k$  clusters. All computations were performed in Matlab (MathWorks, Natick, MA), using its “kmeans( )” function.

An important goal of our cluster analysis was to determine an optimal value for  $k$ . At what value of  $k$  are the feature vectors partitioned into their “natural” clusters, that is, the clusters that correspond to distinctly different color terms? In particular, we wondered whether there were more than the eight chromatic BCTs specified by Berlin and Kay. This is a difficult question to answer with cluster analysis because there is no ground truth against which the solution corresponding to a particular value of  $k$  can be assessed (if there were such a ground truth, it would not be necessary to perform the cluster analysis, and the problem would be merely to classify the terms into their known clusters).

To deal with this issue, we followed an approach based on the *gap statistic* described by Tibshirani, Walther, & Hastie (2001). This approach is based on a comparison between the clustering of the data and the corresponding clustering of a synthesized reference set of feature vectors. The reference set is created by sampling from a uniform distribution of vectors in feature space that have all the statistical properties of the data, except that by design they have no natural clustering. For each  $k$ -means clustering of  $j$  data points into  $k$  clusters, let  $D_i(k)$  be the sum of the distances between points within the  $i$ th cluster ( $i = 1, \dots, k$ ) and the corresponding cluster centroid. Then  $TDD(k)$  is the total dissimilarities across all  $k$  clusters. Now, consider the corresponding clustering of a reference set of  $j$  feature vectors into  $k$  clusters. The gap  $g(k)$  is the difference between  $\log TDD(k)$  and the *mean* of  $\log TDR_n(k)$ , obtained from clusterings of  $n$  sample sets of

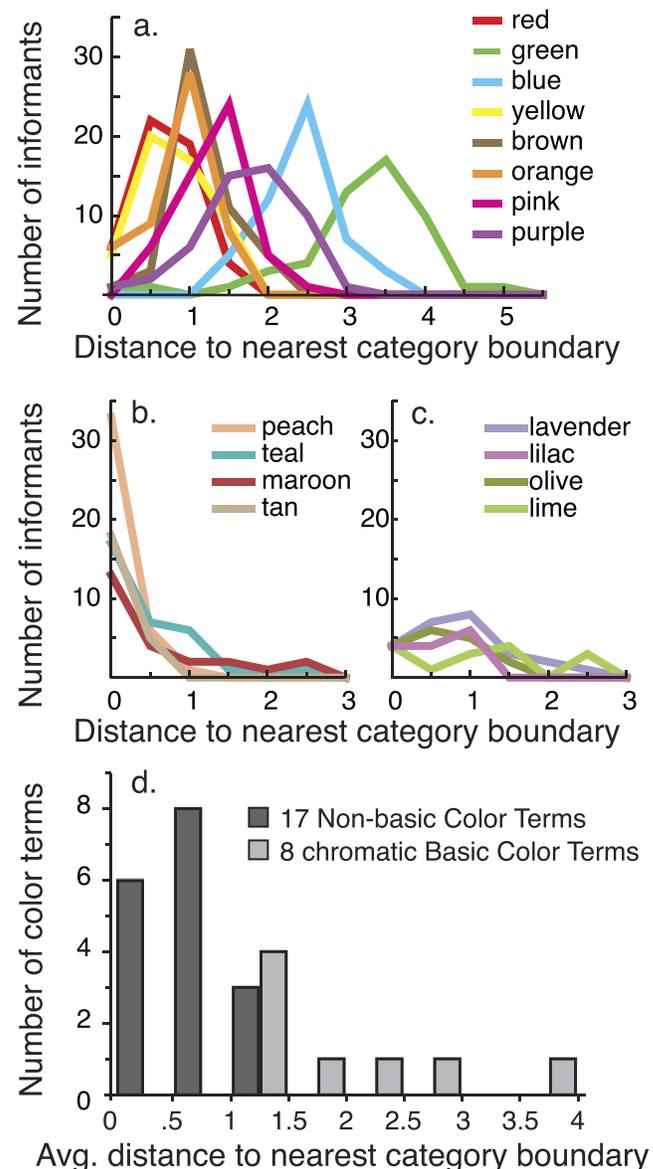


Figure 7. The positions of the centroids of the chromatic color terms, expressed as the unsigned distance (in number of chips) to the nearest BCT boundaries. (a) The BCT centroids. (b–c) The centroids of eight nonbasic color terms. (d) The average locations of the centroids for the 25 chromatic color terms from the free-naming data set on the upper limb in Figure 4.

$j$  feature vectors drawn from the uniform reference distribution:

$$g(k) = \log\left(TDR_n(k)\right) - \log\left(TDD(k)\right) \quad (1)$$

The gap statistic is based on the intuition that if a data set contains exactly  $k_{\text{opt}}$  clusters, then  $g(k)$  will increase for  $k \leq k_{\text{opt}}$ , since  $k$ -means is doing an increasingly better job of reducing within-cluster distances in the data set as  $k$  approaches  $k_{\text{opt}}$  when compared to the distances obtained by clusterings of the reference sets, which by design have no clusters. For  $k > k_{\text{opt}}$ , the partitions must split one or more of the  $k_{\text{opt}}$  clusters, and  $g(k)$  will not continue to improve and may even decline relative to  $g(k_{\text{opt}})$ .

Let  $G(k)$  represent the change in gap between the  $k$ th and the  $k$ th +1 clustering,

$$G(k) = g(k + 1) - g(k) - s_{\text{TDR}}(k + 1), \quad (2)$$

where  $s_{\text{TDR}}(k + 1)$  is an error term related to the standard deviation of the log  $TDR_n(k)$ . Then  $k_{\text{opt}}$  is defined as the largest  $k$  before the first zero crossing of  $G(k)$  (see Tibshirani et al., 2001, for details). Formally, this rule is stated as follows:

$$k_{\text{opt}} = \operatorname{argmax}_k \{k | G(k) > 0\}. \quad (3)$$

Among the virtues of the gap statistic is that it can test for the absence of any clusters in the data set (i.e.,  $k_{\text{opt}} = 1$ ).

Our gap-statistic analysis was based on  $n = 20$  reference sets. To create a reference set, we took the 963 color-naming patterns from our data set. The centroid of each pattern was then randomly relocated to a new location in the coordinate frame of our WCS color-sample space, and a feature vector for this new pattern was created based on the new location. Thus, our reference sets preserved exactly the distributions of the sizes and shapes of regions of color space associated with the color terms observed in the original data set. However, in each reference set, the locations of the centroids of the feature vectors were drawn from a uniform distribution of centroids falling within the coordinate frame of the WCS color chart, and therefore had no natural clusters.

Preliminary studies indicated that despite adopting a best-of-100-replications criterion for the clustering of our data, several independent clusterings of the data for a given value of  $k$  still tended to produce small differences in  $TDD(k)$ . In order to assess the impact of this variation on our gap-statistic analysis, we ran 1,000 separate analyses. For each analysis, we created  $k$ -means clusterings of the data for  $k = 1, \dots, 25$  and compared those to the corresponding clusterings on a new ensemble of 20 reference sets.

Figure 8 shows the values of the gap statistic  $G(k)$  over the 1,000 runs, and the inset shows the distribution

of  $k_{\text{opt}}$ . The minimum value of  $k_{\text{opt}}$  was 11, so there were always at least three more clusters in the data set than the eight chromatic BCTs listed by Berlin and Kay. The mode of the  $k_{\text{opt}}$  distribution (which was close to its median and its mean) at  $k_{\text{opt}} = 17$  was the best estimate of the statistically significant chromatic clusters in the free-naming data set.

Thus, the first cluster analysis of the free-naming data yielded a glossary of 20 color categories (17 chromatic color categories plus *black*, *white*, and *gray*; Table 2). This glossary included nine more chromatic clusters in the free-naming data set than there are BCTs, according to the first conjecture of Berlin and Kay. All informants used more than 11 glossed color categories; the most frequent number was 18 (Figure 3b).

Consensus diagrams of the 17 chromatic color terms appear in Figure 9. In this article, we call them by the names most commonly used by informants (above each diagram in Figure 9; see also Figure 2b). The second result was that even when all the colors were consolidated into their  $k_{\text{opt}}$  clusters, none of the nine statistically significant nonbasic chromatic categories was used by 100% of informants (Table 3b, categories of rank 12–20). The clusters illustrated in Figure 9 were used in the motifs analysis described later in this article.

The glossary derived from the  $k = 17$  solution varied slightly from run to run. Repeated runs of  $k$ -means revealed essentially the same glossary, but occasionally the *lime* cluster shown in Figure 9 was replaced by a *rust* (dark red) category. Most  $k$ -means runs produced cluster centroids that were all confined to contiguous regions of feature space, like those illustrated in Figure 9. Occasionally, however, one of the 17 consensus maps derived from cluster analysis covered two disjoint regions of the color chart. One of the regions was always more prominently represented than the other, and the corresponding cluster centroids were easily associated with one of the observed nonbasic color terms. In any event, the minor run-to-run perturbations in the  $k$ -means derivation of the American English glossary did not affect the main conclusions drawn from the motifs analysis discussed later.

## American English motifs

In their analysis of the WCS data set, Lindsey and Brown (2009) found that the color-naming systems of individual informants around the world fell into about four motifs, and that multiple motifs were present in the data sets of the great majority of the WCS languages. To determine whether these results also apply to American English, we performed a second  $k$ -means analysis on the present data set, based on the glossary of 20 terms revealed by the first  $k$ -means analysis.

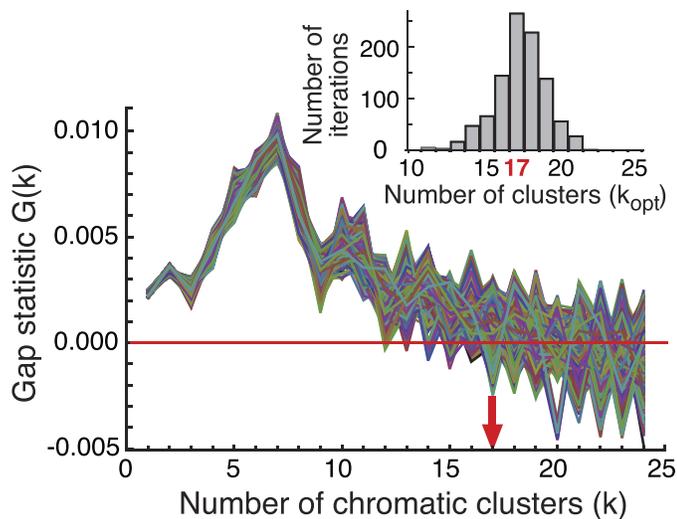


Figure 8. The gap statistic (see Equation 3) as a function of the number  $k$  of chromatic clusters. Each individual line in the main graph represents the results for one of 1,000 comparisons between the  $k$ -means clusterings of the data for  $k = 1, \dots, 25$  and corresponding clusterings of ensembles of 20 reference null sets. See text for details. Inset: histogram of  $k_{\text{opt}}$  derived from the 1,000 functions shown in the main graph. This analysis indicates that there are about 17 chromatic color terms in American English (red arrow in main figure).

For each of the 51 informants, we created a feature vector consisting of 20 elements, one element for each of the glossed terms (the 17 chromatic color terms in Figure 9 plus *black*, *white*, and *gray*; see Tables 2 and 3b). Each of the 20 elements had a value corresponding to the fraction of samples (out of 330) in the WCS chart assigned to that particular gloss by the first cluster analysis. For example, if a given informant named 20 samples using one or more words synonymous with *teal*, then the *teal* element of that informant's feature vector was assigned the value of 0.061 (20/330).

We performed a series of  $k$ -means analyses on the set of 51 informants' feature vectors to create partitions of  $k = 1, \dots, 5$  clusters. There was almost no between-run variation in the  $k$ -means solutions. To determine how many clusters were statistically significant, we created reference feature vectors in ensembles of 20 sets by creating scrambled versions of the informants' feature vectors. We then used the results of the  $k$ -means analysis of the informants' feature vectors and the sets of reference feature vectors to perform a gap-statistical analysis as outlined previously (Tibshirani et al., 2001).

This analysis revealed two statistically significant motifs in American English. Figure 10 shows the consensus maps for the two motifs, as well as the 0.80 threshold consensus maps. The informants whose data fell into the first motif used primarily the eight chromatic

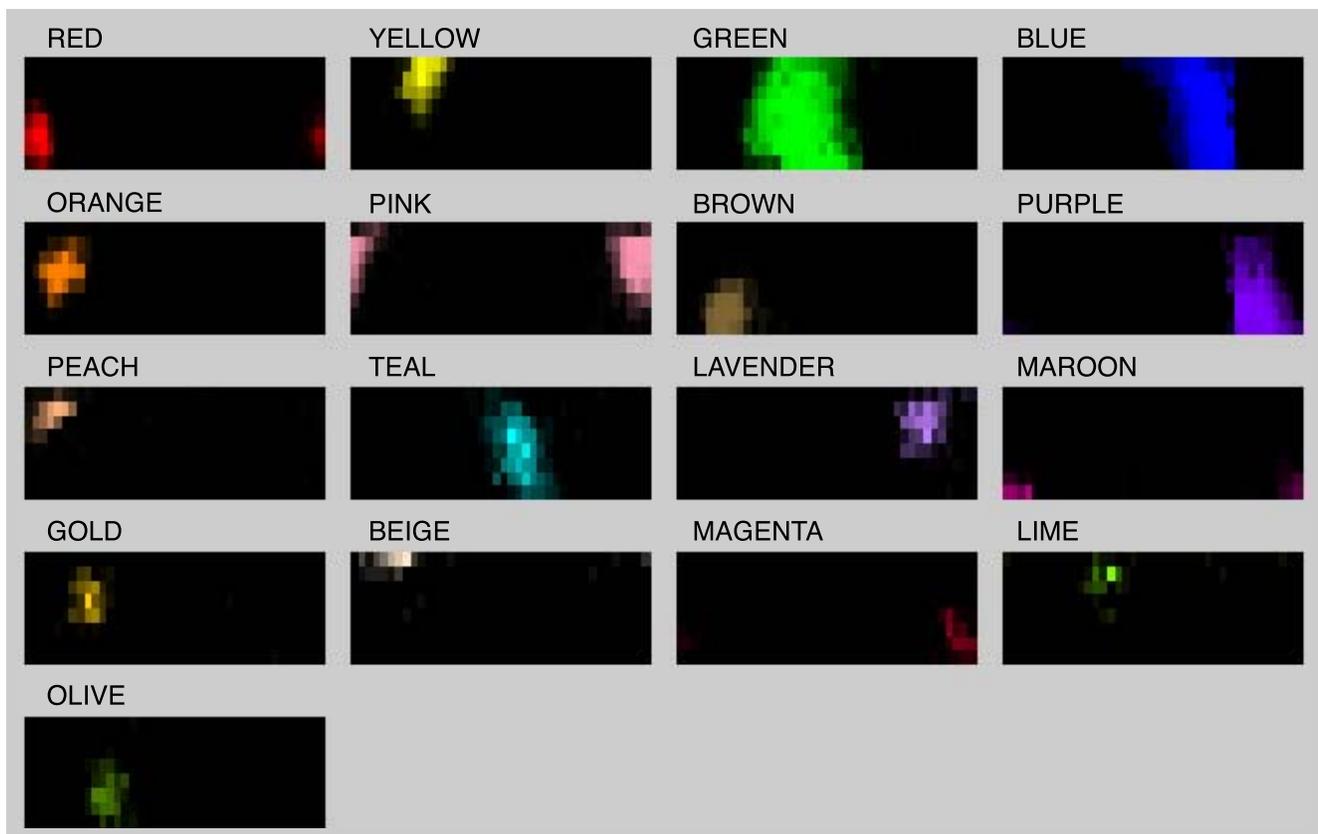


Figure 9. Consensus diagrams of the 17 chromatic color terms identified by the first  $k$ -means cluster analysis.

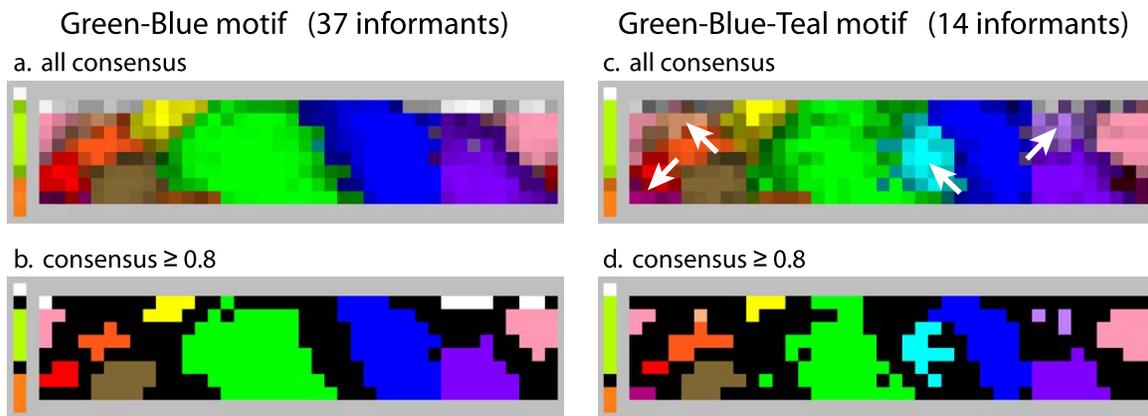


Figure 10. The results of the second cluster analysis, based on the glossed color-naming patterns from the free-naming task. (a–b) Seventy-three percent of 51 informants fell into a cluster that corresponded closely to the green–blue motif of Lindsey and Brown (2009). This motif is similar to the Stage-VII pattern of Berlin and Kay (1969), being composed primarily of the 11 BCTs. (a) Consensus map. (b) 0.8 threshold consensus map. (c–d) The remaining 27% of 51 informants fell into a second, green–teal–blue motif, which is new here. The new motif included the high-consensus nonbasic terms *maroon*, *peach*, *teal*, and *lavender*. These new colors (see arrows) appear in both the consensus map (c) and the 0.8 threshold consensus map (d).

BCTs of Berlin and Kay (plus *black*, *white*, and *gray*). Thus, the first motif is similar to the green–blue (GB) motif in the WCS, which was so-named after the color terms corresponding to the cool colors (Lindsey & Brown, 2009). Informants expressing this motif tended to use the nonbasic color terms in the glossary idiosyncratically and with low frequencies. The informants whose data fell into the second motif also used the 11 BCTs of Berlin and Kay, but they also used some additional terms extensively and consistently, particularly *teal*, *peach*, *lavender*, and *maroon*. Figure 10d shows that consensus for each of these terms equaled or exceeded 0.8 for some of the color samples. We will call this the green–teal–blue (GTB) motif because of the names given to the cool colors. The GTB motif is new, and did not appear in the WCS analysis. Fourteen informants used the GTB motif, whereas 37 informants used the GB motif. Partition of informants' data into the GB and GTB motifs increased color-naming consensus from the overall value 0.74 for the free-naming data set as a whole to an average within-motif consensus of 0.79 for the glossed terms (GB consensus = 0.81, GTB consensus = 0.77). We created 10,000 partitions of the informants' data into two random “motifs” with 37 individuals in one and 14 individuals in the other. The highest average consensus from this simulation was 0.77. Therefore, the statistical significance of our *k*-means-based partition is  $p < 10^{-4}$ .

The results shown in Figure 10 were remarkably robust, as they were essentially independent of the precise representation of informant feature vectors that was chosen, the dissimilarity metric, or even the size of the glossary ( $12 \leq k \leq 22$ ) extracted from the first cluster analysis. For example, in addition to the 20-element informant feature vectors described above, we

performed a separate analysis employing 330 element vectors, where each element was assigned a nominal value representing one of the 20 glossed terms. In this approach, which Lindsey and Brown (2009) used in their analysis of the WCS motifs, the dissimilarity metric was a modified Jaccard coefficient (Leisch, 2006). In yet another version of the motif analysis, the 20-element feature vectors were populated with *z* scores representing deviations of each informant's usage of each glossed term from the mean. The most striking result was that all these approaches agreed very well on the identity of the first two motifs: a green–blue motif and a second motif—green–teal–blue—with high informant usage of *teal*, *peach*, *maroon*, and *lavender*. Beyond two motifs, the various cluster analyses mostly generated minor variations on the green–blue motif and variations of the green–teal–blue motif that emphasized various subsets of the four additional categories in the green–teal–blue motif. The main differences between the various approaches were in their statistical power. The 330-dimension approach revealed only one statistically significant motif, whereas the 20-dimension approach involving *z* scores revealed four statistically significant motifs. The 330-dimension approach from Lindsey and Brown (2009) worked on the WCS data set because of the enormous number of observations, where statistical power was not an issue. However, that approach was apparently underpowered for the present application.

Figure 10c and d also reveals that three of the four high-consensus nonbasic color terms appear in the low-consensus regions of the chart between the BCTs: *peach* appears in the dark, low-consensus area between *pink*, *orange*, *yellow*, and *white*; *teal* appears between *green* and *blue*; and *maroon* appears between *red* and *black*. These color terms evidently name new categories that

arose between categories that previously existed, along the lines proposed by Levinson, and were not the result of partitioning existing color categories into smaller sets. In contrast, the samples called *lavender* in the free-naming task were a proper subset of the *purple* category, consistent with the partition hypothesis of Kay et al.

## The popularity of glossed color terms

The number of informants using each of the terms corresponding to the 20 glossed color categories from the first  $k$ -means cluster analysis (Table 3b) appears as a function of the rank order of their popularity as the upper graph (triangles) in Figure 4. As before, the BCTs were at the ceiling and were fitted with a constant line. There was a clear break in the function fitted to the nonbasic terms no matter how we dealt with ties in rank ordering the data. The four most popular nonbasic terms were fitted with a power law of exponent  $-0.79$  (white triangles), whereas the remaining five terms were best fitted with exponent  $-2.3$  (gray triangles). The four nonbasic categories on the second limb of the function were *teal*, *peach*, *lavender*, and *maroon*, the same terms that appeared in the GTB motif. This clear distinction between the popularity of the four new terms in the GTB motif and the remaining five statistically significant terms provides additional evidence, independent of the second  $k$ -means motifs analysis, that those four additional glossed terms are well integrated into the color lexicons of many informants. It also invites the speculation that the color lexicon of American English is currently undergoing change, and that those four terms are in the process of taking their place along with the original 11 BCTs of Berlin and Kay.

## Gender, age, and color naming

We also examined the free-naming data set to determine whether there was an effect of age and whether the American men and women in this sample differed in the number of terms in their color vocabularies. Figure 11a shows the number of men and women using each of the nonbasic terms used by three or more informants. The terms in Figure 11a were generally used more frequently by women than by men [average difference = 0.070,  $t(39) = 2.50$ ,  $p = 0.017$ , two-tailed]. When the nonbasic color terms were consolidated into categories, this gender difference persisted [Figure 11b; average difference = 0.117,  $t(8) = 2.72$ ,  $p = 0.027$ , two-tailed]. Furthermore, men and women were unevenly distributed across the two motifs (Figure 11c), with women significantly less likely to use the GB motif, and more likely to use the GTB motif, than men:  $t(49) = 2.30$ ,  $p = 0.026$ . On average, men used 9.71 nonbasic

color terms, whereas women used 12.3 nonbasic color terms. A multiple regression of the log-transformed frequency data (to normalize their distribution) against age and gender revealed that this difference between men and women was statistically significant ( $r = 0.304$ ,  $p = 0.0301$ ) but that age was not associated with the number of nonbasic terms ( $r = 0.079$ ,  $p > 0.5$ ). Apparently women's color vocabularies contained more terms than those of men. Women also distinguished more color categories than men did and were more likely to use the motif that contained more color terms. In contrast to the clear effects of gender, this data set showed no statistically significant age effect.

There is a sizable literature on the subject of gender and color naming (e.g., DuBois, 1939; Nowaczyk, 1982; Simpson & Tarrant, 1991; see Biggam, 2012, for a recent review), which generally shows larger color vocabularies among women than among men (but see also Machen, 2002; Sturges & Whitfield, 1995). However, data like these do not indicate whether this difference is biological or social in origin. On the biological side, Jameson, Highnote, and Wasserman (2001) have argued that women identify more color categories than men do because of well-understood sex-linked genetic differences in their long- or middle-wavelength-sensitive (L or M) cone pigments (Nathans, Merbs, Sung, Weitz, & Wang, 1992). Many heterozygous females carry the genes for four types of cone: In addition to the three normal cone pigments, some have the gene for an additional (normal) L cone pigment, and about 10% of females carry the gene for an additional (anomalous) L or M cone pigment. According to Jameson et al. (2001), women who are heterozygous for the two versions of the L cone pigment may divide the spectrum into more color bands than men or women with only three cone-pigment genes. However, very few women who are heterozygous for anomalous trichromacy are actually tetrachromats, in the sense of being able to use the normal and anomalous pigments together to discriminate between colors (Jordan, Deeb, Bosten, & Mollon, 2010), although some apparently experience a subtle influence of their anomalous cones on color appearance under conditions where the influence of the normal cones is minimized. Thus, while a well-documented L-cone gene polymorphism might, in principle, provide a basis for explaining some or all differences between males and females in color naming, the behavioral data obtained from heterozygous women do not provide straightforward, unambiguous support for this explanation. Also on the biological side, there are other biological differences between males and females, for example due to testosterone receptors in the cerebral cortex, that may explain subtle, quantitative differences between men and women in the

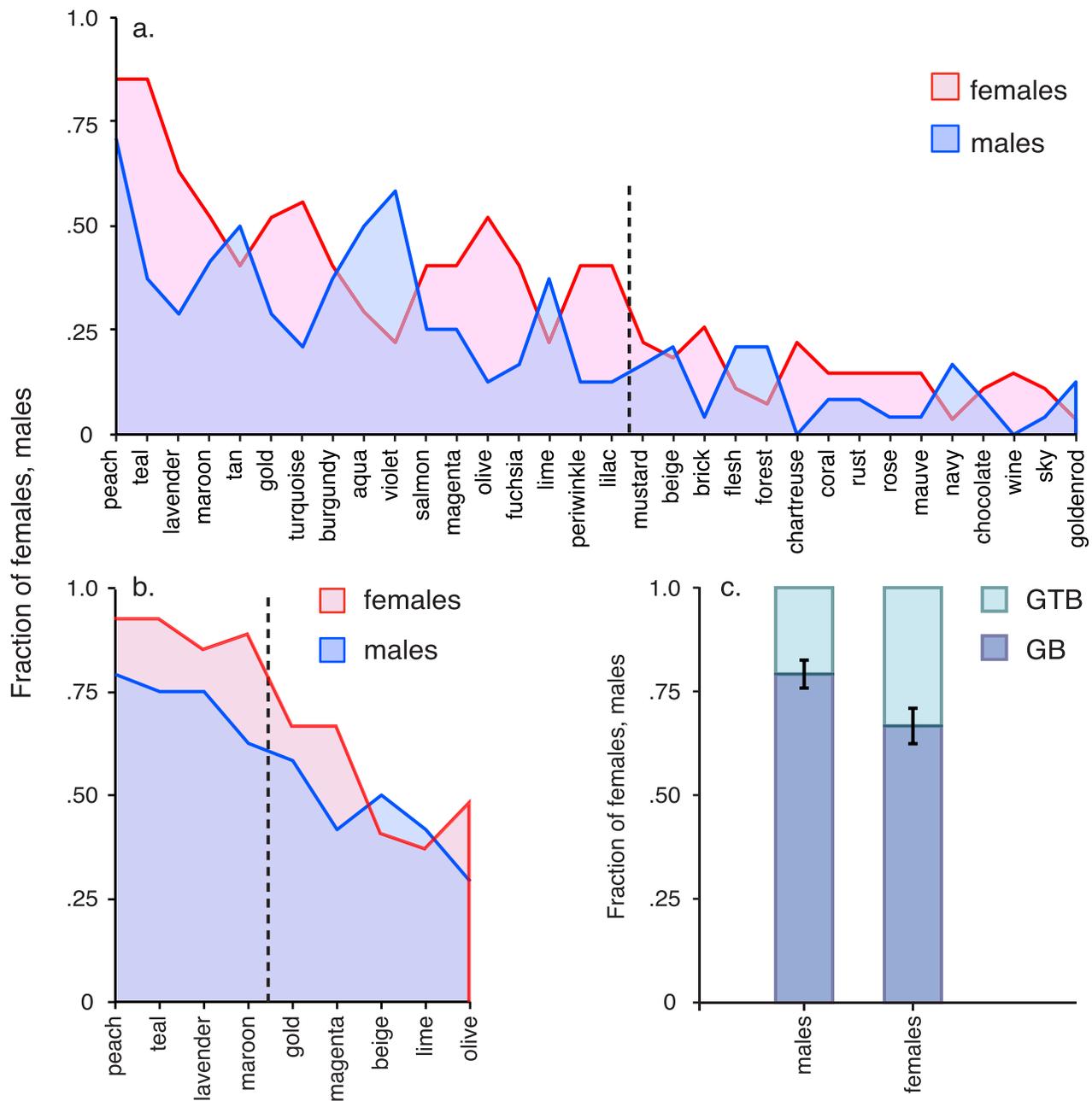


Figure 11. (a) The fraction of men and women who used each nonbasic color term shown in Figure 2a. Vertical dashed line: the break point between the two power-law functions that were fitted to the disks in Figure 4 (Table 2). (b) The pattern of women using more color terms persisted when the nonbasic color terms were consolidated into their corresponding glossed color categories. Vertical dashed line divides the four nonbasic color terms (to the left of the line) in the second motif from the other nonbasic color terms; it is also the break point between the two power-law functions fitted to data in Figure 4 (triangles). (c) The fraction of men and women whose data fell into each of the two motifs. GB: the green–blue motif; GTB: the green–teal–blue motif. Error bars:  $\pm$  one standard error of the dividing line between the two motifs.

appearance of colors (Abramov, Gordon, Feldman, & Chavarga, 2012). The present difference between men and women is probably larger than the subtle sex-related differences of color appearance that were reported by Abramov et al. (2012). These two types of biological difference between males and females might have small combined effects that together exert

a measurable influence on the appearance and naming of colors.

Some investigators have espoused the alternative view that women’s role in society as consumers of the decorative arts has honed their color discrimination into a finer sense of color appearance, resulting in superior color-category naming ability among women

than among men (e.g., Rich, 1977; Swearingen, Layman, & Wilson, 1978). The difficulty with the social hypothesis is that it does not make specific quantitative predictions that could be falsified.

There is evidence in the field of sociolinguistics that language change begins with women and young people, especially in the lower-middle socioeconomic class, with the language of men, older people, and people of other socioeconomic classes changing later (Labov, 1990; Tagliamonte & D’Arcy, 2009). The American men and women in this sample differed in the number of terms in their color vocabularies, which provides further evidence suggesting that the color lexicon of American English is still changing. However, there was no reliable effect of age. Furthermore, data on socioeconomic status were not collected, and our informants probably represented a relatively narrow range along this dimension, so language-change effects related to social class could not be examined.

It is not immediately clear that the gender difference we report here is directly related to color. In addition to the pervasive gender effect outlined by the sociolinguists, Laws (2004) reported other domain-specific differences in vocabulary size between the genders. These approaches suggest that the difference in the size of the color lexicons of men and women might be a specific instance of more general, and less-color-related, phenomena. Of course, it is also possible that subtle biological differences between males and females, combined with the social differences between men and women, are jointly responsible for the reliable gender-related effect. Such a combined explanation would require a model with many free parameters, and it would be even harder to test quantitatively than either explanation alone.

## Discussion

The data for this project were the color names provided for 330 Munsell color samples by 51 native speakers of American English under two instructions: free naming, where any monolexemic color term could be used, and constrained naming, where only the 11 basic color terms (BCTs) of Berlin and Kay (1969) were allowed. This sample of informants used a total of 122 color terms under free-naming instructions, with an average consensus of 0.74, and 11 color terms under constrained-naming instructions, with an average consensus of 0.85. When the free-naming data set was subjected to a cluster analysis similar to that of Lindsey and Brown (2006), 20 statistically significant color categories were discovered: the BCTs of Berlin and Kay plus nine nonbasic chromatic categories. Examination of the centroids of these nine nonbasic

categories revealed that their corresponding color terms were generally deployed to name colors that fell in the low-consensus regions between the chromatic BCTs of the constrained-naming data set, suggesting an “emergence”-like mechanism of color-term evolution. However, two terms, *lime* and *lavender*, were proper subsets that partitioned their corresponding BCTs, *green* and *purple*, suggesting a “partition”-like mechanism.

A second cluster analysis, based on the glossary derived in the first analysis, revealed that this sample of American English-speaking informants expressed two motifs. The first motif, expressed by 73% of informants, was similar to the green–blue (GB) motif observed in the World Color Survey (Lindsey & Brown, 2009). It was also similar to the color lexicon of BCTs listed by Berlin and Kay. The second, green–teal–blue (GTB) motif was expressed by 27% of informants, and included four nonbasic terms that were used with high consensus: *peach*, *teal*, *lavender*, and *maroon*. Women were statistically more likely than men to use the GTB motif.

This prospectively designed study shows that the key features of color naming in the WCS are general to a written language spoken in the United States. These key features are the diversity among the speakers of a single language in how colors are to be named, and the way in which the language’s color terms are deployed across the range of colors that any individual might encounter. Based on the published literature, one might suspect that diversity across individuals in their color vocabularies might be restricted to unwritten languages such as those in the WCS, which are spoken in nonindustrialized societies far from Western influence. Contrary to that supposition, prominent diversity in color vocabulary among individuals who speak American English persisted, even after cluster analysis consolidated the 122 color terms elicited in a free-naming protocol into a glossary of 20 distinct named color categories. This indicates that diversity among informants is common even in American English. One might also suspect that this apparent within-language diversity might be an artifact of each informant’s haphazard color-term choices (from a much larger color idiolect) on the spur of the moment, and might not reflect true individual differences in color cognition. On the contrary, the diversity reported here was not haphazard: Instead, there were two distinct motifs, which repeated themselves with minor variation across the idiolects of these 51 informants. Thus, the phenomenon of multiple within-language motifs observed in the unwritten languages of the WCS also generalizes to at least one written language spoken in an industrialized society, namely American English.

## The first conjecture of Berlin and Kay: The BCTs and their universality

Berlin and Kay's (1969) first conjecture was that every language includes a vocabulary of no more than 11 BCTs, which are distinct from other ordinary color terms that an informant might use. In English, Berlin and Kay's BCTs are *black, white, red, yellow, green, blue, brown, orange, pink, purple, and gray*. According to Berlin and Kay, the BCTs in any language are monolexic, abstract terms that are used by all or nearly all competent speakers of that language, with high consensus and consistency, to name the color of any type of object, including color samples such as those used in the present study. The 11 BCTs of Berlin and Kay were indeed used by all 51 informants here, and they were the only terms that showed such full usage. This general result is certainly consistent with Berlin and Kay's first conjecture.

However, the free-naming data set shows several other features that are less obviously consistent with Berlin and Kay's first conjecture. First, only seven of the BCTs were applied to any samples with 100% consensus, omitting *white, red, yellow, pink, and orange* (Figure 5b), so the requirement that the 11 BCTs show high consensus is not perfectly observed. Second, every informant used at least 12 terms, and the modal number of terms was 18, so the minimum number of terms is greater than 11. A third finding that challenges Berlin and Kay's first conjecture is the frequency-versus-rank power-law functions derived from the popularity data. If the 11 BCTs of Berlin and Kay were the only commonly used terms, and if the nonbasic terms were entirely idiosyncratic in their use, the power-law functions should fall off very steeply for ranks greater than 11. Instead, both unglossed data and glossed data (white disks and triangles, respectively, in Figure 4) fell off with steepness near  $-1.0$  for the first 17 unglossed and the first four glossed nonbasic terms. The expected precipitous decline followed (gray disks and triangles in Figure 4). These results suggest that there are about 28 common terms that are used and understood as part of the core color vocabulary of American English; after glossing, 20 of these terms are statistically significant, and about four of them seemed on their way to frequent use. Those four glossed terms are the key components of the GTB motif, which was expressed by 14 of the 51 informants.

The flow of information from informant to listener is important to understanding the use of color terms in communicating about color. Zipf's law expresses the reciprocal relationship between the frequency of independent observations (e.g., the usage of words in a language, the population of cities; in this case, the popularity of color terms) and the rank order of those frequencies. In the case of language, Zipf's law is

understood to reveal an evolutionary trade-off between the need of the speaker to spend the least effort necessary in communication, at the risk of reduced clarity (producing a small vocabulary), and the need of the hearer for conciseness and clarity in the received message (requiring a larger vocabulary; see Ferrer i Cancho & Solé, 2003, for further discussion). Double-power-law behavior is common and is thought to reflect two processes: one process that governs the creation of a kernel lexicon of limited size consisting of versatile words designed for general but imprecise communication, and the other generating an unlimited lexicon for specific communication (Ferrer i Cancho & Solé, 2001). The unglossed data set shows a break in the function between the regimen for the more popular color terms (ranked 12–28) and the less-popular color terms (ranked after 28). This discontinuity may be an instance of this distinction between common color terms and those chosen on the spur of the moment from a larger color lexicon that each informant has in his or her mind but does not use routinely in everyday communication. Even the glossed data have a steep-slope section, which occurs after four glossed categories. Taken together, the segmented structure of the frequency data suggested that the BCTs are not the whole story when it comes to color naming in American English: Even after consolidating the data into glossed categories, at least four more glossed color categories are commonly used and understood by many informants.

## The second conjecture: Color-term evolution

Berlin and Kay's (1969) second conjecture was that color lexicons evolve over time by adding new color terms. This evolution occurs as societies become technologically more complex, and as distinctions among similar colors become more crucial in the everyday lives of their individual members. We examined this data set to determine whether it provided evidence in favor of the "successive differentiation of existing categories" specifically suggested by Berlin and Kay's partition hypothesis, against the alternative view, which is Levinson's emergence hypothesis (2000), whereby new color terms are added to cover hard-to-name colors that for one reason or another have become particularly salient.

The data set reported here is certainly compatible with the view that the color lexicon of American English is currently evolving. For example, the high popularity of the four most common nonbasic color categories and their inclusion in the second motif suggest that those terms are on their way to joining the BCTs of Berlin and Kay to form a new lexicon of basic terms. However, it is logically impossible for the number of color categories to continue to increase ad

infinitum. Yendrikhovskij (2001) proposed that 16 named color categories is on the high end of basic color-lexicon size, on the basis of his information-theoretic analysis of color terminology. Our Zipf's-law analysis also reveals a steep decrease in color-term popularity beyond 15 glossed terms (triangles in Figure 4). Thus, 15 or 16 terms may be an upper limit on the size of basic color idiolects, at least in the milieu of early-21st-century American English.

The difference between the men and the women in the present sample also suggests that the language related to color is changing, and that women are in the vanguard. However, the lack of a reliable age effect in the present data set suggests that the American English color lexicon is changing slowly compared to the age range of our sample, and it does not suggest that individual informants recapitulate the history of color-lexicon change over their lifetimes. Without historical data collected using consistent methodology, it is not possible to examine these issues definitively.

The results of these analyses provide some evidence for both Levinson's and Kay et al.'s (2010) views of color-term evolution. Much as Levinson suggested, 13 of the 17 frequently used nonbasic color terms (from Figure 4, listed in Table 2) appeared in the low-consensus regions between the BCTs, and their centroids were very near the nearest BCT boundaries (Figures 6 and 7). For example, *peach* appears in the hard-to-name region between *orange*, *pink*, *white*, and *yellow*. In contrast, four of the 17 nonbasic color terms were clearly "partition" colors, as predicted by Kay et al. (2010). For example, *lavender* appeared as a proper subset of *purple*, suggesting that it partitions the large, previously undifferentiated *purple* category into two smaller, more articulately named units. These results generally suggested that both "emergence" and "successive differentiation" probably occur as American English adds new color terms.

Considering Berlin and Kay's evolution conjecture, it is instructive to examine the terms for the light colors. BCTs and nonbasic color terms exist in American English to name the light colors in much of the diagram: *pink*, *peach*, *yellow*, *lime*, and *lavender* (Figures 6d and 12). In contrast, there is no commonly used, high-consensus word that means *light blue* (Table 3a), a finding consistent with Sturges and Whitfield's (1995) report of nonbasic-term usage by informants of British English, and Boynton and Olson's (1987, 1990) studies of American English. *Sky*, the closest term to *light blue*, is included in the unglossed data set in Figures 2b and 12, but it ranks 43rd in popularity in this data set (Table 3a), and does not reach significance in the cluster analysis that created the glosses. In contrast to English, *light blue* has been reported to be a standard (perhaps basic) color term in some Indo-European languages—Russian *goluboj* (Winawer et al.,

2007), modern Greek *ghalazio* (Thierry et al., 2009), (some forms of) Spanish *celeste* (Bolton, 1978, p. 294), and Farsi *asamuni* (Friedl, 1979)—as well as some non-Indo-European languages—(some forms of) Arabic *celesti* (Al-Rasheed et al., 2011; Borg, 2007), and Turkish may distinguish dark blue from light blue, *lacivert* versus *mavi* (Ozgen & Davies, 1998). Of the common light color terms in English, *pink* and *yellow* are BCTs, *peach* is an "emergent" color, and *lime* and *lavender* are "partition" terms. Thus, a *light blue* color term could appear either as a partition of *blue* or as a boundary term between *blue* and *white*. The fact that no common color term, basic or nonbasic, exists for *light blue* suggests that the "emergence" and "partition" mechanisms do not necessarily predict, universally, which color terms will occur. This case study illustrates how little is really understood about how color terms are added to the lexicons of world languages.

There has been considerable speculation over the years about what universal processes guide color-term evolution. Kay and his colleagues have emphasized the importance of universal aspects of the perceptual representation of color appearance; particularly, Kay et al. (2010) emphasized the salience of the Hering fundamental hue sensations of red, green, blue, and yellow, plus black and white, which they supposed to have a well-understood physiological basis. While this account is at least qualitatively in line with modern accounts of human color vision, the physiology underlying the Hering sensations remains obscure (Lindsey & Brown, 2014). Furthermore, this account does not provide any insight into the order in which color terms should appear as color lexicons change (however, see Ratliff, 1976). Recent theories that are based on human perception of color differences predict that color terms should be added in a manner that optimally partitions color space by minimizing color differences within the new contiguous color categories while maximizing color differences between adjacent categories (Jameson & D'Andrade, 1987; Regier, Kay, & Khetarpal, 2007). The simulations by Regier et al. (2007) based on this principle resemble Berlin and Kay's evolution trajectory of color names up to six terms, but the authors do not show simulations for lexicons greater than this number. Therefore, it is not clear that their simulations will generalize to more than six terms. Interestingly, the explanation that Boynton and Olson (1987, 1990) give to explain the possible emergence of *peach* as a new BCT is similar to the optimal partition principle of Regier et al.

Other accounts of color-term evolution have focused on the importance of the statistics of color in the natural environment. Philipona and O'Regan (2006) argued that the *red*, *green*, *blue*, and *yellow* universal color categories extracted from the World Color Survey

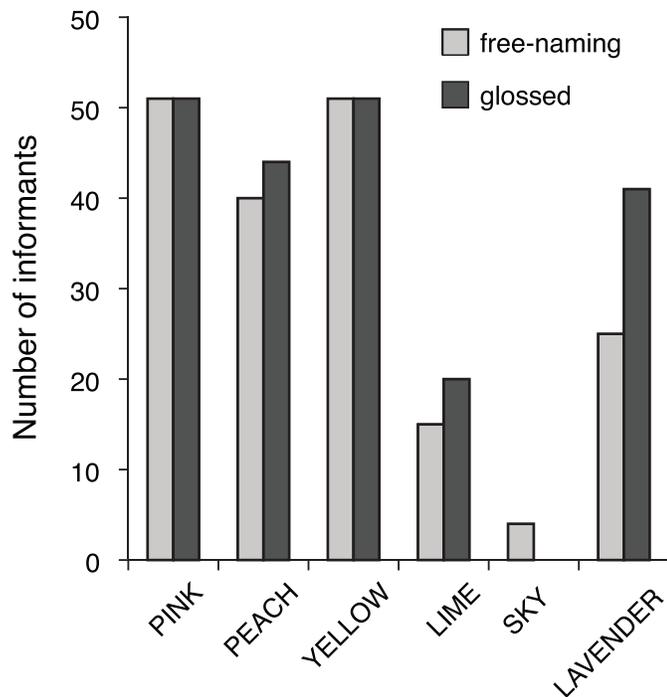


Figure 12. The number of informants using terms for *light red* (pink), *light orange* (peach), *yellow*, *light green* (lime), *light blue* (sky), and *light purple* (lavender), from Figure 2. All but *light blue* are statistically significant color categories in the cluster analysis, and appear in the set of glossed color categories.

are special because the early visual responses to the light from the corresponding surfaces are least perturbed by changes in the environmental illuminant. On the basis of this account, one might expect all the Hering primary-color categories to appear consistently early in color-term evolution. Contrary to that expectation, the WCS data set shows many informants who use *brown*, *gray*, *pink*, or especially *purple* categories, while lacking a *blue* category (Lindsey & Brown, 2009). Yendrikhovskij (2001) has shown that *k*-means clustering ( $k = 2, 3, \dots, 11$ ) of color samples obtained from pictures of natural scenes and represented in the CIELUV uniform chromaticity space recapitulates Berlin and Kay's evolutionary sequence of color-term acquisition. However, Steels and Belpaeme (2005) have shown that Yendrikhovskij's analysis is very sensitive to the perceptual space in which color is represented. Therefore, they suggest that Yendrikhovskij's results are due more to his choice of color model than to the statistics of the natural-scene samples themselves.

Thus there are many deterministic accounts of how human color perception, possibly coupled with color statistics in the natural environment, could guide color-term evolution. The factors that underlie these accounts undoubtedly place important constraints on

color-term evolution: Clearly, not every conceivable color term could come into being, and not every possible motif could evolve at every possible juncture in color-term change. These deterministic accounts resemble theories of biological development in some ways, where change occurs from one state to the next along a single pathway, subject to limited variability across individuals. Contrary to that view, we believe that the trajectory of color-lexicon change is rather more haphazard and evolution-like than is suggested by the more development-like deterministic accounts reviewed previously. It seems more likely that color-lexicon change in general, and particularly the addition of color terms beyond Berlin and Kay's 11 BCTs, is better understood in light of the principles that govern cultural evolution (Henrich, Boyd, & Richerson, 2008). In this view, cultural change is analogous to Darwinian evolution, where change from a given state can occur along any one of many trajectories, subject to a few well-understood principles (see, for example, Xu, Griffiths, & Dowman, 2013). Over time, these principles lead to changes in the relative prevalences of the different "species" (motifs) of color naming within a language community. Consistent with this view, the languages in the WCS differ from one another in the relative frequency of the expressed motifs (Lindsey & Brown, 2009), just as isolated biological populations differ from one another in the fraction of individuals with certain genetic alleles. Several recent studies—Steels and Belpaeme (2005), Komarova, Jameson, & Narens (2007), and Xu et al. (2013), among others—have attempted to model color-term evolution along these lines. Kay and his colleagues are increasingly proposing a less linear and deterministic (and less development-like), and more reticulate and stochastic (and more evolution-like), account of color-term evolution based on the WCS data set. Consistent with this view of color-term change as evolution, the results of the present study have provided clear evidence for both color-naming diversity among speakers of American English and the organization of this diversity into two distinct color-naming motifs.

*Keywords:* color naming, color categories, color appearance

## Acknowledgments

Mr. Kevin Guckes assisted with data collection. We acknowledge A. C. Hurlbert and M. W. Ridley for helpful discussions, and the constructive suggestions of an anonymous reviewer. This project was supported by a Fred Brown Award from the Ohio State University

Department of Psychology, the Ohio Lions Eye Research Foundation, and NSF BCS-1152841.

Commercial relationships: none.

Corresponding author: Delwin T. Lindsey.

Email: lindsey.43@osu.edu.

Address: Delwin Lindsey, PhD, Professor of Psychology, Ohio State University-Mansfield, Mansfield, Ohio, USA.

## References

- Abramov, I., Gordon, J., Feldman, O., & Chavarga, A. (2012). Sex and vision II: Color appearance of monochromatic lights. *Biology of Sex Differences*, 2(1), article 21.
- Al-Rasheed, A., Al-Sharif, H. H., Thabit, M. J., Al-Mohimeed, N. H., & Davies, I. R. L. (2011). Basic colour terms in Arabic. In C. P. Biggam, C. A. Hough, C. J. Kay, & D. R. Simmons (Eds.), *New directions in colour studies* (pp. 53–58). Amsterdam: Benjamins.
- Berlin, B., & Kay, P. (1969). *Basic color terms: Their universality and evolution*. Berkeley, CA: University of California Press.
- Biggam, C. P. (2012). Locating basic colors in the OSA space. *The semantics of colour: a historical approach*. Cambridge, UK: Cambridge University Press.
- Bolton, R. (1978). Black, white and red all over: the riddle of color term salience. *Ethnology*, 17, 287–311.
- Borg, A. (2007). Towards a history and typology of color categorization in colloquial Arabic. In R. E. MacLaury, G. V. Paramei, & D. Dedrick (Eds.), *Anthropology of color* (pp. 263–294). Amsterdam: Benjamins.
- Boynton, R. M. (1997). Insights gained from naming the OSA colors. In C. L. Hardin & L. Maffi (Eds.), *Color categories in thought and language* (pp. 135–148). Cambridge, UK: University of Cambridge.
- Boynton, R. M., & Olson, C. X. (1987). Locating basic colors in the OSA space. *Color Research and Application*, 12, 94–105.
- Boynton, R. M., & Olson, C. X. (1990). Salience of chromatic basic color terms confirmed by three measures. *Vision Research*, 30, 1311–1317.
- Cook, R. S., Kay, P., & Regier, T. (2005). The world color survey database: History and use. In Cohen, Henri, & Claire Lefebvre (Eds.), *Handbook of Categorisation in the Cognitive Sciences* (pp. 224–242). Amsterdam: Elsevier.
- Davies, I., & Corgett, G. (1994). The basic color terms in Russian. *Linguistics*, 32, 65–89.
- DuBois, P. H. (1939). The sex difference on the color-naming test. *American Journal of Psychology*, 52(3), 380–382.
- Ferrer i Cancho, R., & Solé, R. V. (2001). Two regimes in the frequency of words and the origins of complex lexicons: Zipf's law revisited. *Journal of Computational Linguistics*, 8(3), 165–173.
- Ferrer i Cancho, R., & Solé, R. V. (2003). Least effort and the origins of scaling in human language. *Proceedings of the National Academy of Sciences, USA*, 100(3), 788–791.
- Friedl, E. (1979). Colors and culture change in southwest Iran. *Language in Society*, 8(1), 51–68.
- Gladstone, W. E. (1858). *Studies on Homer and the Homeric age (Vol. III)*. London: Oxford University Press.
- Henrich, J., Boyd, R., & Richerson, P. U. (2008). Five misunderstandings about cultural evolution. *Human Nature*, 19, 119–137.
- Jameson, K. A., & Alvarado, N. (2003). Differences in color naming and color salience in Vietnamese and English. *Color Research and Application*, 28(2), 113–138.
- Jameson, K. A., & D'Andrade, R. G. (1987). It's not really red, green, yellow, blue: An inquiry into perceptual color space. In C. L. Hardin & L. Maffi (Eds.), *Color categories in thought and language* (pp. 295–319). Cambridge, UK: Cambridge University Press.
- Jameson, K. A., Highnote, S., & Wasserman, L. (2001). Richer color experience in observers with multiple photopigment opsin genes. *Psychonomic Bulletin and Review*, 8(2), 244–261.
- Jordan, G., Deeb, S. S., Bosten, J. M., & Mollon, J. D. (2010). The dimensionality of color vision in carriers of anomalous trichromacy. *Journal of Vision*, 10(8):12, 1–19, <http://www.journalofvision.org/content/10/8/12>, doi:10.1167/10.8.12. [PubMed] [Article]
- Kay, P., Berlin, B., Maffi, L., Merrifield, W. R., & Cook, R. (2010). *The World Color Survey*. Berkeley, CA: CSLI Publications.
- Kay, P., & McDaniel, K. (1978). The linguistic significance of the meanings of basic color terms. *Language*, 54, 610–646.
- Komarova, N. L., Jameson, K. A., & Narens, L. (2007). Evolutionary models of color categorization

- based on discrimination. *Journal of Mathematical Psychology*, 51(6), 359–382.
- Labov, W. (1990). The intersection of sex and social class in the course of linguistic change. *Language Variation and Change*, 2, 205–254.
- Laws, K. R. (2004). Sex difference in lexical size across semantic categories. *Personality and Individual Differences*, 36, 23–32.
- Leisch, F. (2006). A toolbox for K-centroids cluster analysis. *Computational Statistics and Data Analysis*, 51(2), 526–544.
- Levinson, S. C. (2000). Yéfi Dnye and the theory of basic color terms. *Journal of Linguistic Anthropology*, 10, 3–55.
- Lindsey, D. T., & Brown, A. M. (2006). Universality of color names. *Proceedings of the National Academy of Sciences, USA*, 103, 16608–16613.
- 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, USA*, 206, 19785–19790.
- Lindsey, D. T., & Brown, A. M. (2014). Color appearance, language, and neural coding. In J. S. Werner & L. M. Chalupa (Eds.), *The new visual neurosciences. Vol. 1* (pp. 511–521). Cambridge, MA: MIT Press.
- Lyons, J. (1995). Colour in language. In T. Lamb & J. Bourrieau (Eds.), *Colour: Art & science* (pp. 194–224). Cambridge, UK: University of Cambridge.
- Machen, V. (2002). Color naming by boys and girls. *Perceptual and Motor Skills*, 94, 348–350.
- Mitzenmacher, M. (2003). A brief history of generative models for power law and lognormal distributions. *Internet Mathematics*, 1(2), 226–251.
- Nathans, J., Merbs, S. L., Sung, C.-H., Weitz, C. J., & Wang, Y. (1992). Molecular genetics of human visual pigments. *Annual Review of Genetics*, 26, 403–424.
- Nowaczyk, R. H. (1982). Sex-related differences in the color lexicon. *Language and Speech*, 25, 257–265.
- Ozgen, E., & Davies, I. R. L. (1998). Turkish color terms: Tests of Berlin and Kay's theory of color universals and linguistic relativity. *Linguistics*, 36(5), 919–956.
- Philipona, D. L., & O'Regan, J. K. (2006). Color naming, unique hues, and hue cancellation predicted from singularities in reflection properties. *Visual Neuroscience*, 23, 331–339.
- Ratcliff, F. (1976). On the psychophysical basis of universal color terms. *Proceedings of the American Philosophical Society*, 120(5), 311–330.
- Regier, T., Kay, P., & Khetarpal, N. (2007). Color naming reflects optimal partitions of color space. *Proceedings of the National Academy of Sciences, USA*, 104(4), 1436–1441.
- Rich, E. (1977). Sex-related differences in colour vocabulary. *Language and Speech*, 20, 404–409.
- Rivers, W. H. R. (1901). Vision. In A. C. Haddon (Ed.), *Reports on the Cambridge anthropological expedition of the Torres Straits. Vol. II: Physiology and psychology* (pp. 1–132). Cambridge, UK: Cambridge University Press.
- Saunders, B. M., & Marth, I.-T. (2007). In B. M. Saunders (Ed.), I.-T. Marth (Trans.), *The debate about colour-naming in 19th century German philology*. Leuven, Belgium: Leuven University Press.
- Schontag, R., & Schafer-Priess, B. (2007). Color term research of Hugo Magnus. In R. E. MacLaury, G. V. Paramei, & D. Dedrick (Eds.), *Anthropology of color: Interdisciplinary multilevel modeling* (pp. 107–124). Amsterdam: John Benjamins.
- Simpson, J., & Tarrant, A. W. S. (1991). Sex- and age-related differences in colour vocabulary. *Language and Speech*, 34(1), 57–62.
- Steels, L., & Belpaeme, T. (2005). Coordinating perceptually grounded categories through language: A case study for colour. *Behavioral and Brain Sciences*, 28(4), 469–489.
- Sturges, J., & Whitfield, T. W. A. (1995). Locating basic colours in the Munsell space. *Color Research and Application*, 20, 364–376.
- Sturges, J., & Whitfield, T. W. A. (1997). Salient features of Munsell colour space as a function of monolexic naming and response latencies. *Vision Research*, 37(3), 307–313.
- Swaringen, S., Layman, S., & Wilson, A. (1978). Sex-differences in color naming. *Perceptual and Motor Skills*, 47(1), 440–442.
- Tagliamonte, S. A., & D'Arcy, A. (2009). Peaks beyond phonology: Adolescence, incrementation, and language change. *Language*, 85(1), 58–108.
- Thierry, G., Athanasopoulos, P., Wiggett, A., Dering, B., & Kuipers, J.-R. (2009). Unconscious effects of language-specific terminology on preattentive color perception. *Proceedings of the National Academy of Sciences, USA*, 106(11), 4567–4570.
- 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*, 63(2), 411–423.
- Uchikawa, K., & Boynton, R. M. (1987). Categorical color perception of Japanese observers: Compari-

- son with that of Americans. *Vision Research*, 27(10), 1825–1833.
- Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences, USA*, 104, 7780–7785.
- Xu, J., Griffiths, T. L., & Dowman, M. (2013). Cultural transmission results in convergence towards colour term universals. *Proceedings of the Royal Society, Series B*, 280(1758), 20123073.
- Yendrikhovskij, S. (2001). Computing color categories from statistics of natural images. *Journal of Imaging Science and Technology*, 45, 409–417.
- Zollinger, H. (1984). Why just turquoise? Remarks on the evolution of color terms. *Psychological Research*, 46(4), 403–409.