



ELSEVIER

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Communication efficiency of color naming across languages provides a new framework for the evolution of color terms



Bevil R. Conway^{a,*}, Sivalogeswaran Ratnasingam^a, Julian Jara-Ettinger^b, Richard Futrell^c, Edward Gibson^d

^a Laboratory of Sensorimotor Research, National Eye Institute, NIH, Bethesda, MD, 20892, United States

^b Department of Psychology, Yale University, New Haven, CT, 06520, United States

^c Department of Language Science, University of California, Irvine, CA, United States

^d Department of Brain and Cognitive Sciences, MIT, Cambridge, MA, 02139, United States

ARTICLE INFO

Keywords:

Information theory
Color categories
Communication efficiency
Universal
Cross-cultural

ABSTRACT

Languages vary in their number of color terms. A widely accepted theory proposes that languages evolve, acquiring color terms in a stereotyped sequence. This theory, by Berlin and Kay (BK), is supported by analyzing best exemplars (“focal colors”) of basic color terms in the World Color Survey (WCS) of 110 languages. But the instructions of the WCS were complex and the color chips confounded hue and saturation, which likely impacted focal-color selection. In addition, it is now known that even so-called early-stage languages nonetheless have a complete representation of color distributed across the population. These facts undermine the BK theory. Here we revisit the evolution of color terms using original color-naming data obtained with simple instructions in Tsimane’, an Amazonian culture that has limited contact with industrialized society. We also collected data in Bolivian-Spanish speakers and English speakers. We discovered that information theory analysis of color-naming data was not influenced by color-chip saturation, which motivated a new analysis of the WCS data. Embedded within a universal pattern in which warm colors (reds, oranges) are always communicated more efficiently than cool colors (blues, greens), as languages increase in overall communicative efficiency about color, some colors undergo greater increases in communication efficiency compared to others. Communication efficiency increases first for yellow, then brown, then purple. The present analyses and results provide a new framework for understanding the evolution of color terms: what varies among cultures is not whether colors are seen differently, but the extent to which color is useful.

1. Introduction

It is widely thought that color terms are acquired by all languages in the same order, in a stereotyped sequence determined predominantly by perceptual salience: black and white, then red, green and yellow (either order), blue, brown, purple, pink, orange, and gray (Berlin & Kay, 1969) [reviewed by (Regier, Kay, & Khetarpal, 2007; Zaslavsky, Kemp, Tishby, & Regier, 2019)]. In this scheme, these 11 colors are the complete set of basic color terms. The Berlin-Kay framework is supported by analysis of the best exemplars (“focal colors”) of basic color terms in the color-naming data of the World Color Survey (WCS) of 110 mostly unwritten languages (www1.icsi.berkeley.edu/wcs/data.html) (Kay & Maffi, 1999). The Berlin-Kay framework posits that the colors chosen as best exemplars are universal (Harkness, 1973) and have a physiological origin (Boynton & Olson, 1990), although no

physiological basis has been discovered (Bohon, Hermann, Hansen, & Conway, 2016). Two issues have been raised regarding the WCS data. First, that the color chips in the Munsell array used to obtain the WCS data confound hue and saturation (Lucy & Shweder, 1979) (saturation is the pigment density of a color; hue is the color direction in color space, e.g. red, orange, green etc.). And second, that the task instructions were complex and restrictive (Saunders & van Brakel, 1997).

The covariation of hue and saturation in the Munsell array is potentially problematic because more saturated colors are more salient (Kohlrash, 1935), a phenomenon known as the Helmholtz-Kohlrash effect (Helmholtz, 1867 (1909 Edition)Helmholtz, 1867Helmholtz, 1867 (1909 Edition); Wyszecki & Stiles, 1982). The concern is that the chips identified as focal colors were selected not only because of their hue but also because of their relatively higher saturation which makes them pop out (Lucy & Shweder, 1979; Witzel, Cinotti, & O’Regan,

* Corresponding author.

E-mail address: bevil@nih.gov (B.R. Conway).

2015). This concern is reinforced by several observations: low-saturation stimuli are difficult to categorize (MacLaury, 2007; Olkkonen, Witzel, Hansen, & Gegenfurtner, 2010; Shamey, Zubair, & Cheema, 2019); focal-color probability is correlated with saturation (chroma) of Munsell chips (Witzel et al., 2015); and participants will pick as best exemplars the highest saturation stimuli among chips of the same hue (Paramei, D'Orsi, & Menegaz, 2014).

The covariation of hue and saturation in the standard Munsell array is reflected in the irregular shape of the Munsell space. Given this geometry, for a pre-defined number of categories, optimal partitioning will determine the boundaries of the categories (Jameson & D'Andrade, 1997; Regier et al., 2007). Optimal partitioning also predicts the evolutionary sequence of color-category acquisition (Zaslavsky, Kemp, Regier, & Tishby, 2018). But the validity of these analyses depends on the accuracy of the geometry of perceptual color space. Unfortunately the underlying principles that give rise to color-space geometry—of Munsell space or any color space—are not well understood. The Munsell geometry was probably sculpted not only by perceptual factors but also by the task and materials. In a departure from Ewald Hering's conception of color, which is defined by four so-called unique hues (red, green, blue, and yellow (Hering, 1905)), Albert Munsell predefined five chromatic anchor points (red, yellow, green, blue, purple). It is not clear why Munsell chose to include purple, or to exclude other basic colors such as orange. One possible contributing factor is that Munsell was limited by materials: he could only work with the pigments at hand, even if he might otherwise have desired additional colors. Regardless of the underlying reasons why Munsell chose five chromatic anchor points, and those specific five, the choices likely had an impact on the geometry of the space because holding explicit category assignments can amplify perceptual biases (Bae, Olkkonen, Allred, & Flombaum, 2015). It seems likely that the five chromatic colors in Munsell's space are more salient not because of hard-wired properties of perceptual color space, but because the task Munsell set for himself injected a bias.

It is often thought that color space is objectively determined by perception. The fallacy of this assumption is belied by the diversity of color spaces in use today (e.g., CIELAB, Munsell, NCS), each with its own geometry. Such diversity shows that task instructions and materials inevitably play a substantial role in determining the geometry of color space (Kuehni & Schwartz, 2008). Color spaces are perhaps better thought of as inventions, not a discoveries. The upshot is that the explanatory power of any account of color categories that depends on a definition of perceptual space is weakened simply because the geometry of color space is ill-constrained. Saturation is the least well-defined parameter of color, and variation in how saturation is measured or represented accounts for much of the geometric variation among color spaces. These considerations underscore the potential utility of a metric for assessing color categorization that is not confounded by saturation.

The second issue that complicates interpretation of WCS data relates to the task instructions (Saunders & van Brakel, 1997). The instructions required that color-naming responses be consistent among participants, evident in all idiolects, monolexemic, abstract (i.e. not used to refer to specific objects), and not borrowed from other languages. The complexity of these instructions likely left room for the people conducting the experiments to bias or coach participants (Saunders & van Brakel, 1997). Moreover, the complexity raised the possibility that the different teams who collected the data implemented different versions of the experiment (Gibson et al., 2017). The validity of the instructions is further undermined by anthropological work showing that color terms invariably originate with object names (Levinson, 2000), such as the term *orange* which derives from the fruit; moreover, many languages borrow color terms from other languages. Relaxing the constraints on participant responses increases the glossary of high-consensus color terms (Lindsey & Brown, 2014), and supports the idea that color-category evolution is less stereotyped than the framework originally proposed by Berlin and Kay (Haynie & Bower, 2016; Lindsey & Brown,

2014).

Besides the issues related to task instructions and the confound of hue and saturation, there is yet another problem with the Berlin-Kay framework: it is not compatible with empirical data obtained by two independent groups showing that languages at so-called early stages in their color-term evolution nonetheless have a complete representation of color distributed across the population: Lindsey, Brown, Brainard, and Apicella (2015) in the Hadza of Sub-Saharan Africa; and Gibson et al. (2017) in the Tsimane' of the Amazon delta. The Berlin-Kay theory stipulates that early-stage languages are not capable of categorizing some colors, whereas the Lindsey and Gibson results show that complete color-categorization knowledge is evident in the population even if most individuals within the population are not capable of categorizing all colors. Taken together, the cumulative evidence underscores the need for an alternative to the Berlin-Kay framework for thinking about color-term evolution.

Rather than using focal-color assignments, a promising alternative approach to understand color categorization behavior leverages information theory. This method uncovers the efficiency with which people communicate about color by measuring inter-subject variability in color naming (Gibson et al., 2017; Lindsey et al., 2015; Lucy & Shweder, 1979; Regier, Kemp, & Kay, 2015; Steels & Belpaeme, 2005; Stefflre, Vales, & Morley, 1966). Consider a communication game. Two people each have access to the same array of color chips. One person picks a specific chip in the array and uses a color word to describe it. How many guesses does it take the listener to figure out exactly which chip was selected given the word? The answer is a metric of informativity: color chips associated with fewer guesses are more informative and have lower surprisal, a term coined by Tribus (Tribus, 1961). Instead of thinking of color systems in terms of categories and their best exemplars, the information theoretic approach thinks of color systems in terms of how efficiently they communicate different color percepts. In our implementation, participants are asked to label each colored chip using a term that they think another speaker of their language would understand—there are no restrictions on the terms that can be used (Gibson et al., 2017).

Here we pick up the use of information theory to understand differences in color-naming across languages. As languages increase in overall communication efficiency about color, do some colors undergo relatively greater increases in communication efficiency? One simple hypothesis is that if the Berlin-Kay theory of color-term evolution is correct, then gains in communicative efficiency should follow the trajectory of color terms in the Berlin-Kay evolutionary sequence. The rich WCS data might provide a way to address this question if the issues regarding task complexity and the confound of hue and saturation can be overcome. We previously showed that communication efficiency estimates from a restrictive task like the WCS are comparable to those obtained using a free-labeling task (Gibson et al., 2017), showing that the complex instructions of the WCS do not impact information theory analysis. Here we address the second issue: are communication efficiency scores for color naming influenced by color-chip saturation? We addressed this question using original data collected in the Tsimane' people of the Amazon basin (Leonard et al., 2015). Like language groups included in the WCS, the Tsimane' have had limited exposure to industrialized society. Industrialization has a dramatic impact on visual diet: almost every scene in industrialized society contains synthetically colored surfaces. It is plausible that industrialization impacts perception, cognition, and naming of color. Thus data obtained from the Tsimane' allow an assessment of the impact of saturation on color naming without the potential confound of industrialization. We found that communication-efficiency estimates were not correlated with saturation, unlike focal-color selections. These results validate further information theory analysis of the WCS data. The analysis of the WCS data presented here shows relative shifts in communication efficiency among colors as languages undergo overall increases in communication efficiency about color. The patterns suggest a new framework for how

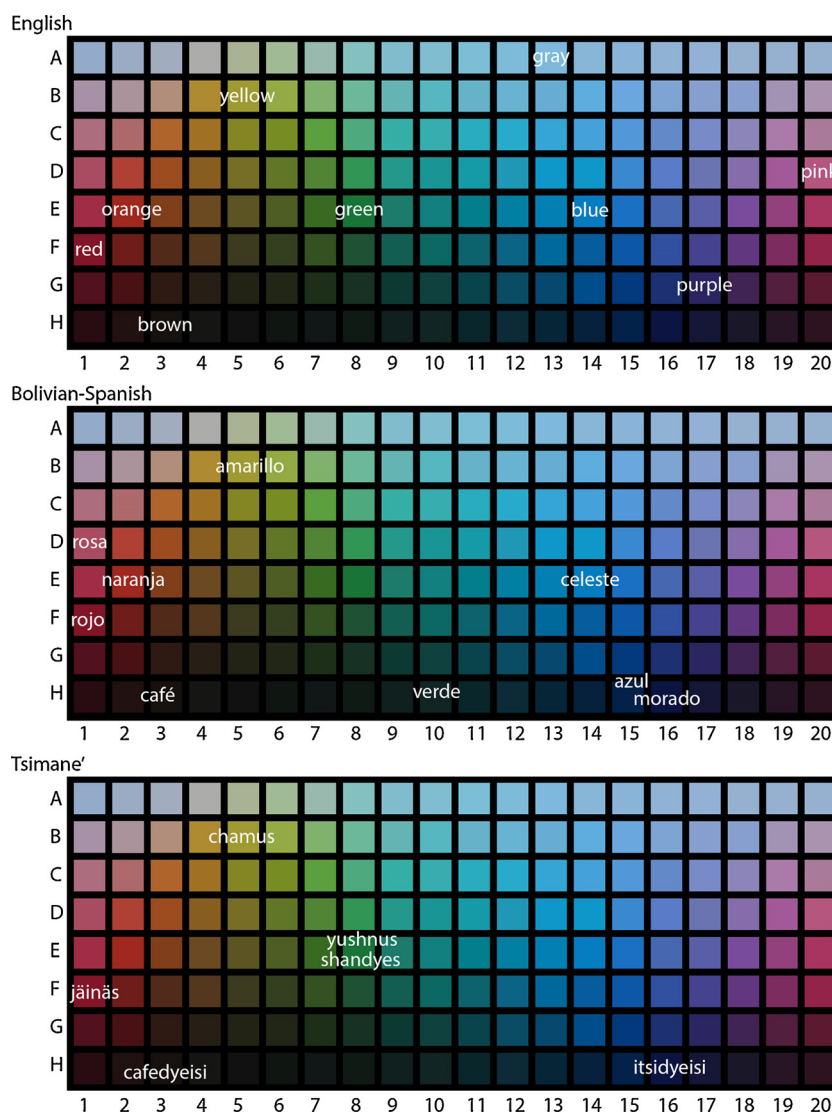


Fig. 1. Focal colors in English, Bolivian Spanish, and Tsimane' identified in an open-response paradigm. Participants were free to label colors using any terms they thought would be understood by someone who spoke their language. We defined the most frequently used modal terms as the set of focal color terms in the language. See Table 1.

color-naming systems evolve.

2. Materials and methods

2.1. Participants

Original data were collected from three language groups: Tsimane' spoken by the indigenous Tsimane' people of the Amazon; Bolivian-Spanish spoken by people in towns neighboring the Tsimane'; and American English (throughout this report, all speakers of "English" spoke American English). Data in the Tsimane' are important because the Tsimane' have little exposure to industrialized society and therefore have little experience with artificially colored objects. Thus the Tsimane' are comparable to the other language groups in the WCS insofar as their cultural history minimizes confounds on color naming brought about by industrialization. Some of the data we obtained in these groups were reported in Gibson et al. (2017); the data and analyses presented here have not previously been reported. There were two experiments, as described more fully below: a color-labeling experiment (participants were shown a single color chip at a time and asked to label it), and a focal-color identification experiment (participants were shown an array of 160 color patches and asked to pick out best

exemplars of various color terms). Fifty-eight Tsimane'-speaking adults (mean age: 33.2 y; SD: 12.8 y; range: 16–78; 38 females); 20 Bolivian-Spanish-speaking adults (mean age: 29.0 y; SD: 9.1 years; range: 18–55; 11 females); and 31 English-speaking adults (mean age: 37.1 y; SD: 11.6 years; range: 21–58; 10 females) completed the color-labeling task (described in Section 2.3). The number of participants for the focal-color task was 99 Tsimane' (mean age 34.1 y; SD: 14.4 years; range: 16–76; 62 female); 55 Bolivian Spanish (mean age 27.6 y; SD: 9.7 years; range: 18–51; 30 female); 29 English (mean age 26 y; SD: 8.9 years; range: 18–55; 14 female).

2.2. Materials

The materials for the color-labeling task consisted of 80 colored chips evenly sampling the standard 320-chip Munsell array of colors used in the WCS (Munsell Book of Colors, glossy collection). Each color chip was about 2 cm square mounted on white cardboard. Materials for the focal-color identification task consisted of an array of 160 color chips evenly sampling the skin of the Munsell color space. The array of colors used in the focal-identification task was organized in an 8 × 20 grid, mounted on matte black cardboard, and each color was a square about 1 cm across, separated from other colored squares by ~3 mm.

We indexed the colors A–H according to lightness, and 1–20 according to hue.

2.3. Procedure

Before performing any color-naming tasks, each participant took a test of normal color vision (Neitz & Neitz, 2001). All participants that failed this task (~ 5% of participants) were excluded from further study. All experimental procedures were approved by MIT's Committee on the Use of Humans as Experimental Subjects; all methods were performed in accordance with the relevant guidelines and regulations. Informed consent was obtained from all participants, as required by the Committee.

For the color-labeling task, participants in the three language groups were presented with each of 80 colored chips evenly sampling the standard 320-chip Munsell array of colors, and simply asked to label each colored chip using a term that they thought another speaker of their language would understand. Note that our experimental approach avoids the complex instructions associated with the World Color Survey (i.e. we did not have to pre-define Basic Color Terms). The chips were presented in a different random order for each participant under controlled lighting conditions using a light box (9 phosphor broadband D50 color-viewing system, model PDV-e, GTI Graphic Technology, Inc. Newburgh, NY).

Following the color-naming experiment, the Tsimane' and Bolivian-Spanish participants were then presented with a uniform sampling of the standard Munsell array of colors (160 of the 320 chips in the standard array, illuminated by the lightbox), and were asked to point out the best example of the color terms that they themselves had produced in the color-naming experiment (Fig. 1).

From these data for Bolivian-Spanish speaking and Tsimane'-speaking participants, we identified for each color chip the modal term used, and defined the set of basic terms as the most commonly used modal terms. We then conducted the focal-labeling experiment on an additional set of Tsimane' and Bolivian-Spanish participants, asking each participant to identify the best examples of the empirically defined basic colors. Most Bolivian-Spanish speakers were asked about *rojo*, *verde*, *amarillo*, *azul*, *celeste*, *anaranjado*, *morado*, *cafe*, and *rosa*. For most Tsimane' participants, we asked about *jainäs* (~red), *yushñus* (~blue, green, or grue), *shandyes* (~green, blue, or grue), *itsidyeyisi* (~purple), *cafedyeyisi* (~brown), and *chamus* (~yellow). A few Tsimane' and Bolivian-Spanish participants did not spontaneously use the modal color terms when doing the initial color-naming experiment. Three Bolivian Spanish participants did not spontaneously use *celeste* in the initial color-naming experiment; and several Tsimane' did not use *itsidyeyisi*, *cafedyeyisi*, *chamus*, *yushñus*, and/or *shandyes*. We did not ask participants to identify focal colors for terms they did not use. English speakers were asked about red, green, yellow, blue, orange, brown, purple, and pink.

2.4. Data analysis

2.4.1. Focal-color analysis

To determine the focal colors, we pooled data across both groups of participants: those from the initial color-labeling experiment in whom we queried best exemplars for the most frequent color terms provided by each participant (thus the terms queried could vary from participant to participant); and those additional participants in whom we queried focal colors for the most frequently used color terms recovered for the population of participants that completed the initial color-labeling experiment. These data establish the likelihood that any given chip would be assigned a basic term, and the probability across the array that any given chip would be assigned any of the basic terms (total participants for focal-color analysis: 99 Tsimane'; 55 Bolivian Spanish; 29 English) (see Fig. 2). Pooling across the two experimental groups is justified because participants in the first color-naming experiment typically used

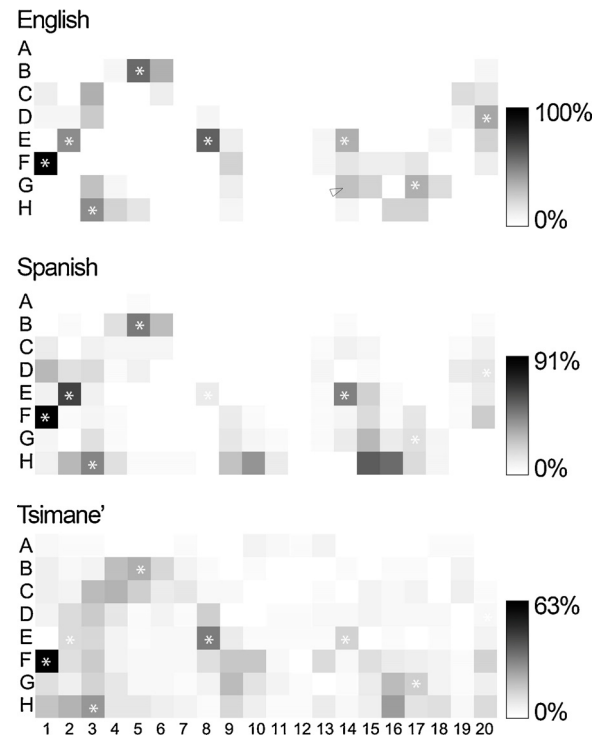


Fig. 2. Probability of being selected as a focal color, regardless of the color term queried (grayscale shows % of respondents in each language group; N = 99 Tsimane'; 55 Bolivian Spanish; 29 English). Asterisks identify the same set of chips to facilitate comparison across the languages. The pattern of probabilities is broadly similar across the three language groups. Note that Bolivian Spanish has a strong peak at H15 (*azul*) and at E14 (*celeste*), indicating that the “blue” part of color space is carved up into two focal colors. The data suggest that English “blue” may also be bimodal with a dominant peak at E14 and a minor peak at G14 (open arrow).

the empirically defined basic terms. Trials in which a Tsimane' or Bolivian Spanish participant did not use a particular basic term were excluded from the focal-color analysis for that term. The number of participants studied exceeds the number used in most color-naming tasks; an analysis of split-halves of the data produces the same conclusions, showing that the sample sizes are sufficiently large. 95% C.I. are reported for all analyses.

2.4.2. Information-theoretic analysis of color-naming data

We computed communication efficiency estimates for the data we collected in Tsimane', Bolivian Spanish, and English, as well as for the data of the 110 languages of the World Color Survey (WCS). Individual trial data of the WCS are available on the internet (<https://www1.icsi.berkeley.edu/wcs/data.html>). Communication efficiency was defined as in Eq. (1), and can be described by a communication game (Baddeley & Attewell, 2009; Lantz & Steffire, 1964; Steels & Belpaeme, 2005) in which a speaker uses a color term to describe a color chip in the array, and a listener has to guess the chip given the word. Chips associated with more guesses have correspondingly higher average surprisal (S , measured in bits). The average surprisal score for each color c is computed by summing together a score for each word w that might have been used to label c , which is calculated by multiplying $P(w|c)$ by $-\log(P(c|w))$, the listener's surprisal that w would label c . We estimate $P(c|w)$ via Bayes Theorem.

$$S(c) = \sum_{\omega} P(\omega|c) \log \frac{1}{P(c|\omega)} \quad (1)$$

The analysis assumes a uniform prior of the probability of a color given a word $P(c)$. The uniform prior is consistent with previous work

on color naming (Gibson et al., 2017; Lindsey et al., 2015; Regier et al., 2007), and fits how we presented the color chips to participants (and how the chips were presented to WCS participants): a participant was initially shown the grid of 160 colors evenly spaced over the saturated Munsell color chips, and they were told they would be asked to label colors from that grid in a randomly presented order. Results using a non-uniform prior that reflects the salience of the stimuli yields similar results (see SI material from Gibson et al., 2017).

2.4.3. Relative shifts in communication efficiency among focal colors

The relative shifts in communication efficiency among focal colors were computed using the WCS data. The WCS data consist of responses from individual participants from 110 languages, in which each participant was asked to label the color of each chip in the standard array of Munsell colors. From the individual responses in each language to each color chip, we estimated a normal distribution of the average surprisal values (see above for the methods used to compute the surprisal values). We analyzed these normal distributions for the seven color chips identified in English as the best exemplars of red, orange, yellow, green, blue, brown, and purple. We randomly sampled a surprisal value from each of these seven distributions, for all languages, and performed binary k-means clustering. The result of the classification analysis binned each of the seven colors into either the low-communication efficiency cluster or the high-communication efficiency cluster. This sampling procedure was repeated 100 times: on each iteration we randomly drew (with replacement) from the normal distributions of surprisal values for each language, for each color chip. This bootstrapping procedure provides an estimate of the percentage of times the color chip ended up in each cluster. The classification of the seven colors for each language were rank ordered according to the overall communication efficiency of the languages. We report the 95% C.I. of the clustering.

3. Results

Color names were queried from speakers of Tsimane', an indigenous people of the Amazon delta that has had limited contact with industrialized society (Leonard et al., 2015). For comparison we also obtained data in two industrialized populations, Bolivian Spanish, and English. In the first part of the study, we present the results on focal colors in the three languages in which we obtained original color-naming data. We show that focal color probability is correlated with color-chip saturation. This finding supports prior work and underscores the risk of using focal-color probability as a metric for evaluating how color-naming systems evolve. We then discovered that average surprisal values (obtained from information theory analysis) were not correlated with color-chip saturation, providing a better metric of color-labeling behavior. In the second part, we use information theory analyses applied to the languages of the WCS to reveal a pattern of color-term evolution among languages.

3.1. Focal color analyses

Bolivian Spanish and English both have relatively well-developed color-naming systems according to the Berlin/Kay evolutionary scheme. Tsimane', meanwhile, has a relatively simple color system: they have only 3 or 4 high-consensus color terms (Gibson et al., 2017). From among the 160 chips in the color array, the chips most often selected as focal colors for the terms queried in the three language groups are given in Table 1. The consistency with which subjects selected a given chip as the focal color for a given color term varied across the languages and colors, ranging from 100% of English subjects who selected F1 as the "red" chip, to only 14% of Tsimane' subjects who selected E8 as the "yushnus" chip. The location of the focal colors among the three languages were roughly comparable, allowing us to gloss the Bolivian-Spanish and Tsimane' terms with likely English translations

(Table 1): In Tsimane': *jäänäs* (~red), *yushñus* (~blue, or green, or "grue"), *shandyes* (~green, or blue, or "grue"), *itsidyeyisi* (~purple), *cafedyeyisi* (~brown), and *chamus* (~yellow); In Bolivian-Spanish: *rojo* (~red), *verde* (~green), *amarillo* (~yellow), *azul* (~blue), *celeste* (~light blue), *anaranjado* (~orange), *morado* (~purple), *café* (~brown), and *rosa* (~pink).

The location of the best exemplars of color terms within the color array was consistent across the three groups, supporting conclusions of the WCS regarding the location of focal colors (Fig. 1). Red, yellow, and brown were in precisely the same location of the color array, in all three languages; orange was in the same location in English and Bolivian Spanish (a term corresponding to orange was not recovered as a basic color term in Tsimane'); pink was within one chip in English and Bolivian Spanish (pink was not recovered as a basic color term in Tsimane'). The term glossed as purple in Bolivian Spanish and Tsimane' was associated with a darker and bluer chip than the chip identified as *purple* in English. The consistency in location of the focal color terms cannot be attributed to investigator bias because we did not pre-define basic color terms. Instead the terms we queried were estimated empirically, as the most frequently recovered set of modal terms elicited in our task. The methodological differences between our experiment and the WCS precluded us from querying some basic color terms, such as a term for orange in Tsimane'. But the methodological differences also allowed us to include some terms that would have been excluded by the WCS instructions such as *cafedyeyisi* (the term *cafedyeyisi* would have been prohibited because it derives from the Bolivian Spanish word "café").

Despite the general consistency in which chips were identified as focal colors across the three languages, there was some variability, which was greatest in the blue/green regions of the color space. Individual Tsimane' participants often used *yushnus* and *shandyes* to reliably indicate two different regions of the color array, but some participants would label what in English we call "green" with *yushnus* while other participants would label this region *shandyes*. The most frequently chosen focal color chip for *yushnus* was identical to that for *shandyes* (E8), but this chip was only selected 14% of the time as *yushnus*, and 17% of the time as *shandyes*. Ambiguity in the blue/green part of color space has been documented for other languages, giving rise to the notion that some languages have a single color category "grue" that encompasses what in English comprises two distinct categories, green and blue (Lindsey & Brown, 2004). Reanalysis of the original color-naming data of the WCS suggests that speakers of "grue" languages do not have a single color category for the blue/green region, but rather experience two color categories within this part of color space, but with low color-naming consensus in the population (Lindsey et al., 2015; Regier & Kay, 2004). Our results on the Tsimane' are consistent with this idea, showing that the Tsimane' have two terms that carve up the blue/green region into two categories, but many Tsimane' are unclear about which term applies to which of the two categories. Ambiguity in the blue/green region also manifests in differences in the way the two post-industrialized languages (English and Bolivian-Spanish) treat this region. Bolivian-Spanish, but not English (or Tsimane'), identified two modal terms for the region called blue in English: *celeste* (for light blue) and *azul* (for dark blue). A similar subdivision of "blue" is also found in other Spanish-speaking cultures (Harkness, 1973; Lillo et al., 2018). Future work is needed to address how naming patterns for the blue/green part of color space evolve. For example, it is not known among the Tsimane' whether *shandyes* and *yushnus* are synonyms, or if these represent distinct color categories as predicted by Regier and Kay (2004).

Fig. 2 shows heat maps of the probability that a chip would be identified as a focal color regardless of the color term that was queried. These maps uncover a similar pattern across the three languages, which is consistent with the general observation that focal colors are found in more-or-less the same locations within the color space across languages, regardless of the overall sophistication of the language's color-naming

Table 1
Focal color chips.

Language	Color	Focal chip	Munsell code	Proportion choosing this color	N
English	blue	E14	10B5/12	0.31	29
English	brown	H3	5YR2/4	0.45	29
English	green	E8	10GY5/12	0.62	29
English	grey	A13	5B9/2	0.31	29
English	orange	E2	10R5/16	0.45	29
English	pink	D20	10RP6/12	0.34	29
English	purple / violet	G17	5P3/10	0.31	29
English	red	F1	5R4/14	1.00	29
English	yellow	B5	5Y8/14	0.59	29
Bolivian Spanish	azul (~ blue)	H15	5PB2/8	0.56	55
Bolivian Spanish	café (~ brown)	H3	5YR2/4	0.44	55
Bolivian Spanish	celeste (~ light blue)	E14	10B5/12	0.48	52
Bolivian Spanish	verde (~ green)	H10	10G2/6	0.38	55
Bolivian Spanish	naranja (~ orange)	E2	10R5/16	0.65	55
Bolivian Spanish	rosa (~ pink)	D1	5R6/12	0.25	55
Bolivian Spanish	morado (~ purple)	H16	10PB2/10	0.51	55
Bolivian Spanish	rojo (~ red)	F1	5R4/14	0.91	55
Bolivian Spanish	amarillo (~ yellow)	B5	5Y8/14	0.47	55
Tsimane'	jainäs (~ red)	F1	5R4/14	0.63	99
Tsimane'	yushnus (~ blue, green, or "grue")	E8	10GY5/12	0.14	99
Tsimane'	shandyes (~ green, blue, or "grue")	E8	10GY5/12	0.17	99
Tsimane'	itsideyeisi (~ purple)	H16	10PB2/10	0.27	90
Tsimane'	cafedyeisi (~ brown)	H3	5YR2/4	0.24	93
Tsimane'	chamus (~ yellow)	B5	5Y8/14	0.18	91

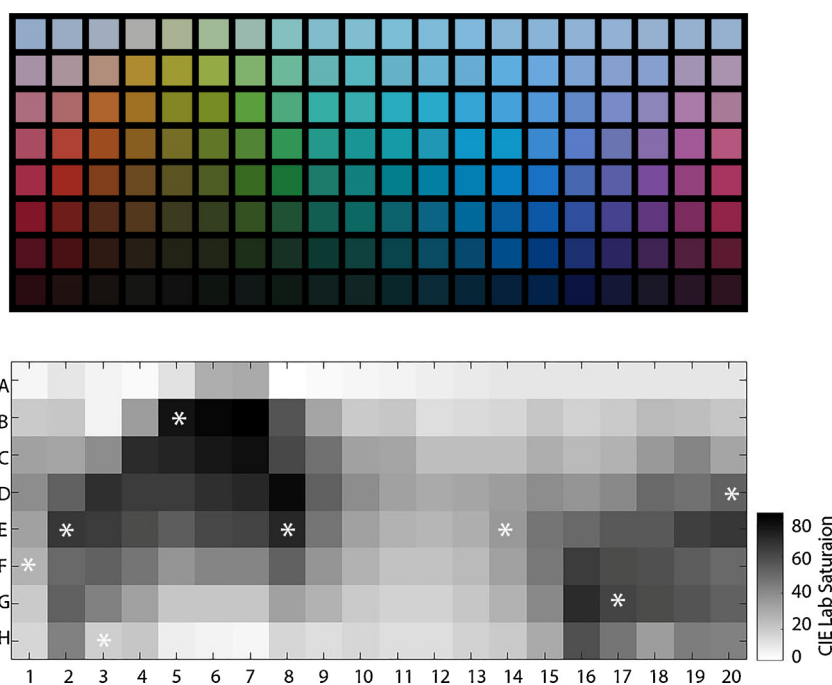


Fig. 3. The standard Munsell array (top) used in our study and the saturation of the color chips in array (bottom). Asterisks show the most frequently selected best exemplars of the English basic color terms *red, orange, brown, yellow, green, blue, purple, pink*.

system (Lindsey et al., 2015). The asterisks in Fig. 2 identify the same chips—the focal colors in English—across the three languages. The location of the focal colors are comparable to prior results given the sampling limits of the array we used. For example, focal blue identified here (10B5/12) is the closest color chip in our array to the focal blue chip identified in other studies (2.5PB5/12) that used a larger array (Sturges & Whitfield, 1995). Interestingly, the results presented in Fig. 2 suggest the existence of two peaks in the blue region of English (arrow head and asterisk in column 14, top panel, Fig. 2), which might indicate that English is in the early stages of carving up “blue” into two basic categories like Bolivian Spanish. Confirming previous work (Lucy &

Shweder, 1979; Witzel et al., 2015), the focal chips were correlated with the most saturated chips in the array (Fig. 3).

To quantify the observation that focal colors were more saturated than other colors in the array, we plot for each chip the probability that the chip will be named a focal color versus its saturation (Fig. 4A), using CIE chroma of the Munsell chips. Fig. 4 focusses on data for Tsimane’ because this group has little contact with industrialized society, minimizing the likelihood that the relationships are impacted by experience with artificially colored objects. Supplementary Fig. 1 shows data for all three languages (Tsimane’, English, and Bolivian Spanish). Focal color probability was correlated with saturation in all three

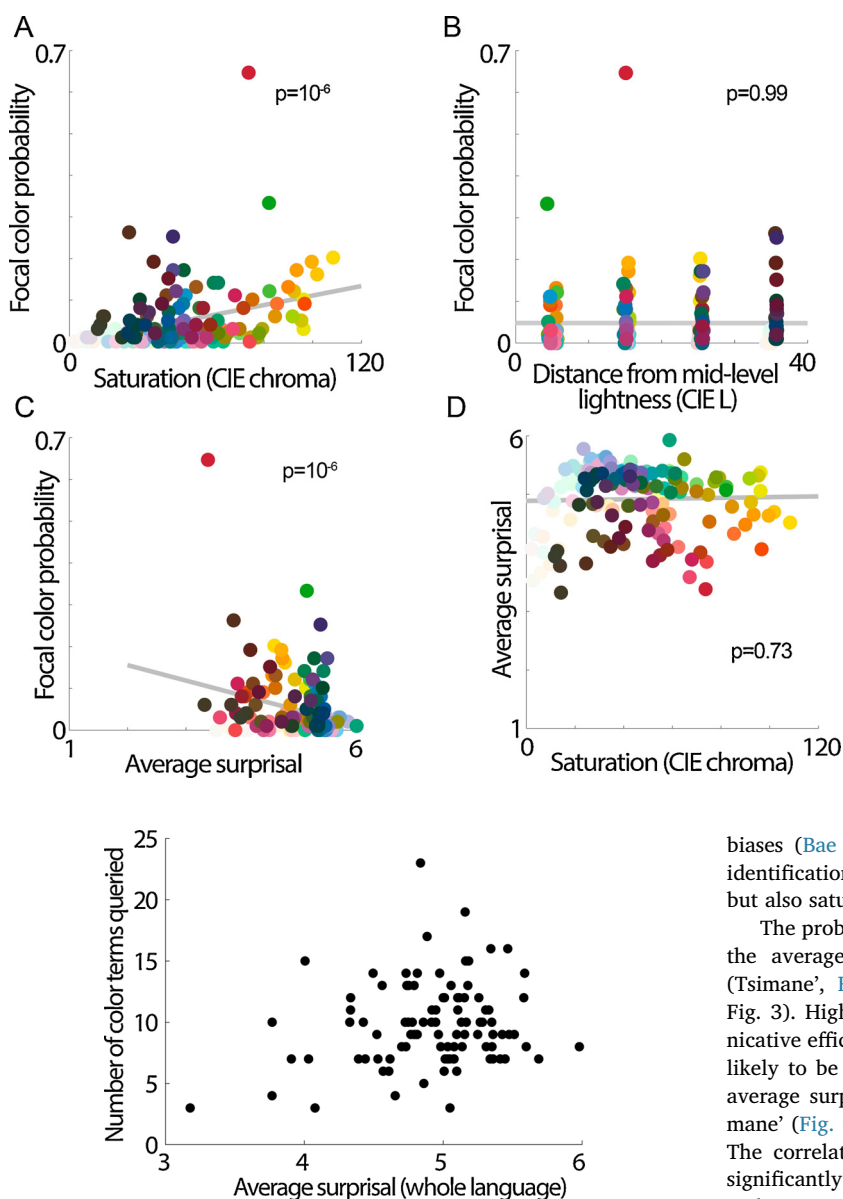


Fig. 5. The average surprisal of a language as a function of the number of focal color terms queried in the language (data from the World Color Survey), $\rho = 0.1$ ($p = 0.3$).

groups; the Tsimane' showed the strongest correlation ($r = 0.36$), comparable to the correlation obtained for the WCS ($r = 0.41$ reported by (Witzel et al., 2015)).

One reviewer asked what impact the lightness of the chip has on focal color probability. Figure 4B plots focal color probability versus the absolute value of the difference between the Munsell chip lightness and the average lightness across chips (Fig. 4B; Supplementary Fig. 2 shows the plots for all three languages). The plot shows no correlation.

The data in Fig. 4 were obtained using a simpler task compared to the WCS. We can therefore conclude that the tendency for best exemplars of high-consensus terms to have relatively higher saturation generalizes from conditions where color categories are pre-defined by the experiment (as in the WCS) to a situation in which the best exemplars are empirically determined for each language. This is an important observation because it could have been the case that pre-defining basic color categories influences how participants weigh the impact of saturation, an alternative suggested by work showing that tasks that involve explicit category assignments can amplify perceptual

Fig. 4. The relationship between focal-color status, lightness, color saturation, and communication efficiency. Data are for Tsimane'. Similar results were obtained for English and Bolivian Bolivian Spanish (see Supplementary Fig. 2). **A.** Focal color status is correlated with the saturation of the color chips. $\rho = 0.36$ ($p = 3 \times 10^{-6}$). **B.** Focal color status is not correlated with stimulus lightness ($p = 0.99$; the CIE Lightness values are absolute values of the difference between the CIE lightness value for each chip and the average CIE lightness values among the chips). **C.** Focal color status is correlated with the average surprisal of the chips, showing that focal color chips are communicated more efficiently. $\rho = -0.28$ ($p = 4 \times 10^{-4}$). **D.** The average surprisal of the color chips is not correlated with saturation. $\rho = 0.027$ ($p = 0.73$).

biases (Bae et al., 2015). Taken together, the results show that the identification of a chip as “focal” reflects not only the importance of hue but also saturation in shaping how we categorize the color chips.

The probability of being a focal color was inversely correlated with the average surprisal of the chips in all three language groups (Tsimane', Fig. 4C; the two industrialized cultures, Supplementary Fig. 3). Higher average surprisal values correspond to lower communicative efficiency, thus the results in Fig. 4C show that the chips more likely to be identified as focal were also easier to communicate. But average surprisal values were not correlated with saturation, in Tsimane' (Fig. 4D), Bolivian Spanish, or English (Supplementary Fig. 4). The correlation between focal color probability and saturation was significantly different from the correlation between average surprisal and saturation in the data from Tsimane' (ANCOVA $p = 0.0004$). This correlation was not significant in the data obtained in English ($p = 0.12$) or Bolivian Spanish ($p = 0.33$); we had fewer participants in these two groups, and extensive experience with industrialized goods could be confounding factors. We also found no correlation between communicative efficiency and the inferred number of basic color categories within a language (Fig. 5). These results show that in non-industrialized populations average surprisal provides a way of assessing color categorization behavior that is not confounded by saturation, and justify the use of the WCS data to examine relative shifts in communication efficiency among colors.

3.2. Relative shifts in communication efficiency among colors

Studies of color-naming evolution have not tracked how color-naming patterns in any given language change over many generations, because it is not obvious how one would obtain such historical data for most languages. Thus to address the question, we assume that each language sampled in the WCS provides a snapshot of the underlying path along which color-naming systems evolve. The evolutionary pattern is estimated by arranging the languages in terms of their overall communicative efficiency about color—the most advanced color-naming systems are assumed to have the highest overall

communication efficiency. Our approach is substantially different from other assessments of color-term evolution, such as by Berlin and Kay, in that we consider the fundamental unit of analysis to be not the color term, but rather the efficiency with which a given color percept is communicated by the color-naming system as a whole. Recent work has used information-theoretic arguments to explain and reproduce the Berlin & Kay categories (Zaslavsky et al., 2018); our goal is to uncover new results and generalizations about color systems that come from thinking entirely in terms of communicative efficiency.

The communication efficiency framework is an alternative to the Berlin-Kay approach for thinking about color-term evolution. Nonetheless, can we make predictions about how communication efficiency scores among colors should shift as languages evolve if the Berlin-Kay evolutionary framework is correct? Our simple prediction is that improvements in communication efficiency would follow the Berlin/Kay color-term evolution sequence. But it is not possible to make a quantitative prediction without making substantial (and unfounded) assumptions about what parts of the color space remain un-named at any given evolutionary stage and how the addition of a new term impacts communication efficiency estimates of already acquired terms. Thus it may be better to treat the communication-efficiency analysis in the present work and the focal-color analysis that is the basis of the Berlin-Kay framework as independent ways of thinking about color-term variability.

Fig. 6 shows how the average surprisal values for each of the eight focal color chips (red, orange, yellow, pink, blue, green, purple, brown) changes as a function of the overall communicative efficiency of the language—languages on the left in each panel are assumed to be at an early stage in color-naming development, while languages on the right are at a more advanced stage of development. The eight chips used in the analysis shown in Fig. 6 were the same for all languages, enabling a direct comparison across languages; the chips were the most frequently chosen focal colors identified in our sample of English-speaking participants. The use of the same chips across languages is justified because the focal colors are very similar across languages (Abbott, Griffiths, & Regier, 2016; Kay & Regier, 2003; Lindsey & Brown, 2006). The panels in Fig. 6 are the same, but isolate, top to bottom, the 95% confidence intervals for yellow, brown, pink, and purple. The average surprisal values for all colors decrease as the languages evolve, but the communication efficiency improvements are greater for some colors than others. The vertical dashed lines were drawn at the points where the average surprisal of the color term became consistently different from blue (“consistent” was defined by a span of 20 languages; brown showed two crossing points). Supplementary Fig. 5 shows the same analysis in which the average surprisal score for each focal color was computed by pooling the scores across all chips selected across the English-speaking population for each focal color, rather than using the average surprisal for the single most frequently chosen chip for each focal color. The number of chips selected across the population for a given focal color ranged from 1 (for red) to 9 (for pink) (Supplementary Fig. 6). The pattern of results is consistent between Fig. 6 and Supplementary Fig. 5.

Fig. 6 uncovers three results. First, as languages increase in communicative efficiency, red and orange remain yoked in their communicative efficiency as the most well-communicated colors, while blue and green remain yoked in their communicative efficiency as the least well-communicated colors. Second, yellow, brown, and purple show the greatest change in average surprisal (pink is discussed below). Third, the point along the x-axis at which yellow, brown, and purple drop in average surprisal differs: yellow drops first, followed by brown, then purple. The vertical dashed lines in Fig. 6 show where the average surprisal for each color becomes consistently significantly different from the average surprisal for blue (there are two transitions for brown).

To quantify the results, we trained a binary classifier to bin the surprisal values across languages for the seven color chips identified as

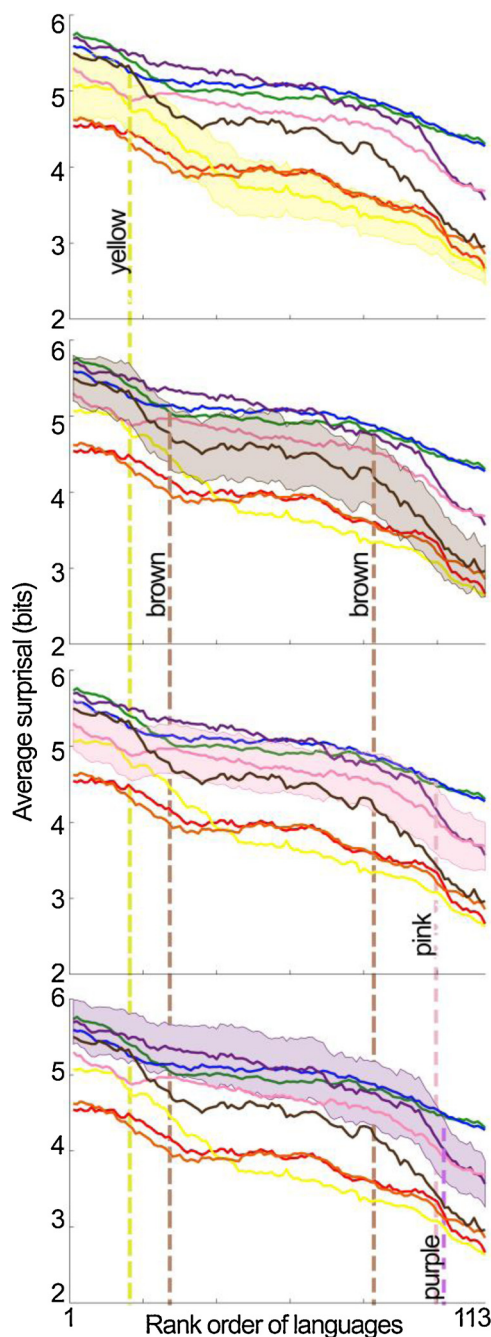


Fig. 6. Relative change in communication efficiency of colors as a function of the overall communicative efficiency of the color-naming system. As languages increase in communicative efficiency about color, blue and green remain the least efficiently communicated colors and red and orange remain the most efficiently communicated colors. Yellow, brown, purple and pink show relatively higher shifts in communication efficiency. Brown undergoes an initial drop in average surprisal, coinciding with the change in communicative efficiency of yellow, and a second drop in average surprisal coinciding with the improvement in communication efficiency of purple/pink. Each panel shows the same data, with the 95% confidence intervals of yellow, brown, pink, and purple (top to bottom). Vertical dashed lines show the point at which the average surprisal for the focal colors are significantly different from blue (95% C.I.). The 95% C.I. were generated by bootstrapping the individual responses in the WCS data for each language.

focal colors in English for red, green, blue, yellow, orange, brown, and purple (see Methods; Fig. 7). The two bins correspond to high communication efficiency and low communication efficiency. Through

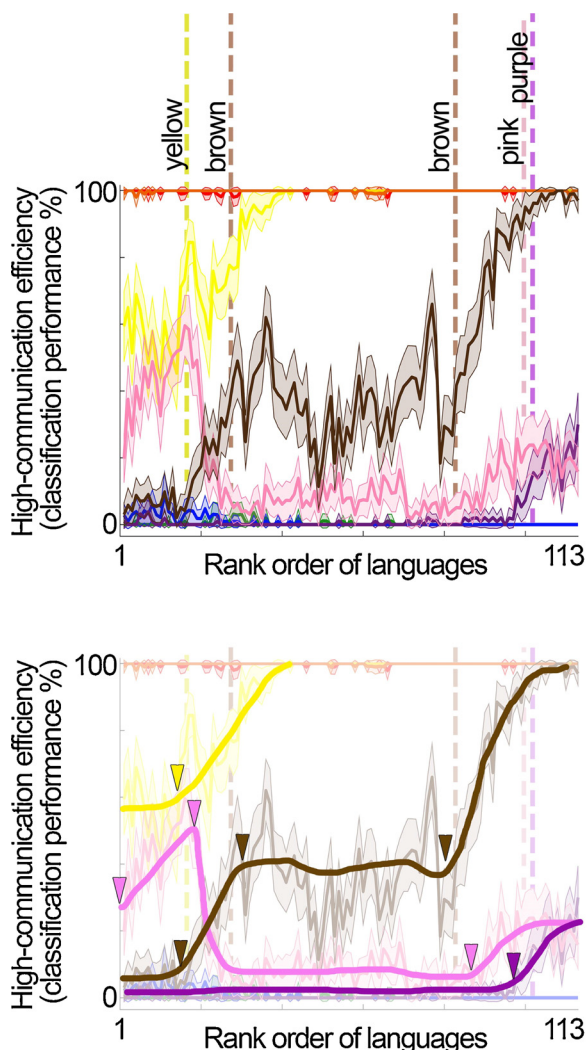


Fig. 7. Binary classification of surprisal values for focal-color chips in each language of the World Color Survey; 95% C.I. shown. **A.** From the individual responses in each language to each color, we estimated a normal distribution of the surprisal values for the color chips identified in English as focal red, orange, yellow, green, blue, brown, and purple. We randomly sampled a surprisal value from each of the seven distributions for all languages and performed binary k-means clustering, into either a low-communication efficiency bin or the high-communication efficiency bin. The sampling procedure was repeated 100 times, providing an estimate of the percentage of times the color chip ended up in each bin. The classification of the seven colors for each language were rank-ordered according to the overall communication efficiency of the languages. Dashed vertical lines correspond to the point along the x-axis where the focal colors differed from blue (as shown in Fig. 6). **B.** Smooth curves with arrowheads indicating the inflection points for the data shown in (A). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

bootstrapping, we obtained probability estimates that each focal color, in each language, would be classified into one or the other bins. The results in Fig. 7 show these probability estimates on the y-axis, rank ordered along the x-axis according to the overall communication efficiency of each language. For all languages, the blue and green chips are always classified into the low-communication bin, and the orange and red chips are classified into the high-communication bin. But within this universal pattern, the yellow, pink, brown, and purple chips shift between low communication and high communication, and they do so at different points along the x-axis. The vertical dashed lines facilitate comparison with Fig. 6 and correspond to the positions along the x-axis where the surprisal values were significantly different from blue. Arrow

heads in Fig. 7B show inflection points in the major trend lines of the classification rates. Consistent with the results in Fig. 6, yellow is the first color to show a consistent increase in communication efficiency, followed by brown, and then pink and purple. A relative increase in communication efficiency is apparent for pink among languages of the lowest communication efficiency, but this trend reverses for languages of intermediate communication efficiency and is recovered again languages with high overall communication efficiency.

4. Discussion

Here we obtain new knowledge into how color-naming systems evolve, using an information theory analysis of color-naming data. The analysis provides a measure of the communication efficiency of color terms: how efficiently can a listener identify a color selected by a speaker given the term used by the speaker? By analyzing communication-efficiency for colors in Tsimane', Bolivian Spanish, English, and across the 110 languages of the World Color Survey, we discovered that as languages progress towards greater overall communication efficiency about color, some colors undergo more dramatic shifts in communicative efficiency. These relative shifts are embedded within a universal pattern in which warm colors are always distinguished from cool colors by having higher communicative efficiency (Gibson et al., 2017). The results provide an alternative to the Berlin-Kay framework for color-term evolution.

The new framework can be captured by three stages, with each stage adding a new color to those that can be distinguished with high communication efficiency (Fig. 8). Stage I: red and orange; Stage II: red, orange, yellow, and partially brown; Stage III: red, orange, yellow, brown, and partially purple. Blue and green remain poorly communicated even in the final stage. For languages of highest communicative efficiency, the colors cluster into three groups: warm (red, orange, yellow, brown), cool (blue, green), and intermediate (purple, pink). Although it is difficult to make a quantitative prediction for how communication efficiency scores among colors should shift if the Berlin-Kay framework is correct (see Results), the present results appear to support a framework that is different from the Berlin-Kay scheme, in which one might predict red to show the highest communication efficiency, followed by green and yellow, then blue, brown, purple, pink, orange, and gray. Two examples illustrate the distinction between the new framework and the Berlin-Kay framework. First, in the present scheme, high communication efficiency about orange emerges early (i.e. among languages with overall low color communication

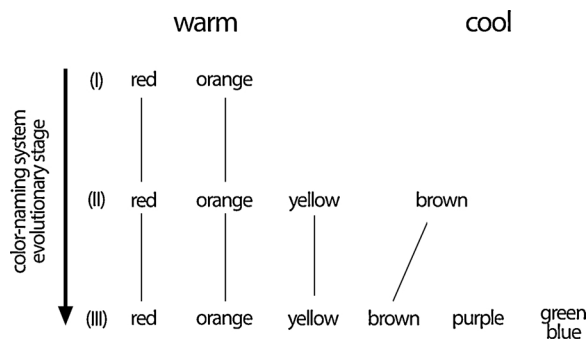


Fig. 8. Stages in the evolution of color language suggested by the information theory analysis of color-naming data across the World Color Survey. The stages are determined from an analysis of the relative efficiency of naming colors, as a function of the overall increase in communication efficiency about color (see Fig. 5). All languages, regardless of their color-naming systems, show higher communicative efficiency for red and orange, compared to green and blue. As the overall communicate efficiency of a color-naming system increases, some colors undergo relatively greater improvement. Stage I has consensus categories for red, orange, and “cool”; Stage II: red, orange, yellow/brown, and “cool”; Stage III, red, orange, yellow/brown, purple and “cool”.

efficiency), whereas in the Berlin-Kay scheme, *orange* emerges late (i.e. among languages at an advanced stage of color-naming evolution). Second, the communication-efficiency analysis does not distinguish between blue and green for any language, whereas the focal-color analysis reveals distinct blue and green categories even for languages at relatively early stages of color-term evolution (Regier & Kay, 2004).

The stages of the new framework do not constitute a stereotyped and finite sequence of color terms, which is a substantial departure from the BK framework. We think the present framework is closer to a true account of how color terms evolve for two reasons. First, languages are continually evolving, even languages such as English which have many high-consensus color terms. For example, Lindsey and Brown review evidence that *cream* and *peach* are emerging as basic color terms in English (Lindsey & Brown, 2014). The continued evolution of color naming, evident in a language at the pinnacle of the Berlin-Kay hierarchy, suggests that the stereotyped sequence proposed by Berlin and Kay is either incomplete or inaccurate. Second, languages do not show a monotonic progression of color-term evolution—they can lose and acquire color terms multiple times (Haynie & Bower, 2016). Taken together with these other observations, the present results suggest the need for a paradigm shift away from the Berlin-Kay framework in how we think about color-term evolution, not a need to substitute one stereotyped sequence for another.

We do not dispute the key observation made by Berlin and Kay: within the Munsell array, languages with a comparable number of high-consensus terms carve up the space in a similar way (Lindsey & Brown, 2006; Regier, Kay, & Cook, 2005; Gibson et al., 2017); and optimal partitioning of Munsell space, given a set number of categories, predicts how languages categorize colors of the Munsell space (Regier et al., 2007; Zaslavsky et al., 2018). Rather, we dispute the explanation that these observations are evidence for a universal, neurobiologically constrained basis for color categorization. The focal-color analysis that is the basis for the Berlin-Kay framework depends upon asymmetric properties of Munsell color space (Jameson & D'Andrade, 1997; Regier et al., 2007). The problem, as we see it, is that these asymmetries are not hard-wired properties of perceptual color space. Instead the peculiar geometry of Munsell color space, like the geometry of any color space, reflects task demands used to create the space. Two properties of the Munsell space likely account for the Berlin-Kay framework: first, the color chips confound hue and saturation; and second, the space is built on the assumption that color depends on five chromatic anchor points. The Berlin-Kay framework can be explained by coupling these features with tradeoffs between many precise color categories and few coarse categories. Our central argument is that a valid theory of color naming must not depend on peculiarities of the color space used to collect the data. Because the Berlin-Kay framework reflects the confound of hue and saturation in the space (Figs. 3 and 4) it is untenable as an accurate, objective account of color naming. The Berlin-Kay framework fails because the Munsell system is not an objective reflection of perceptual color space (see Introduction)—indeed, it seems unlikely that any representation of color space is complete, objective, and universal. It remains unknown whether there is a connection between color perception and color categorization (Witzel & Gegenfurtner, 2018), probably because color perception itself has many dimensions. For example detection of color targets at threshold saturation levels does not tell us much about how colors appear above threshold, yet both phenomena are considered "perception". The multitude of color spaces in use today reflect prejudices about what which aspects of perception are most relevant.

The Munsell space was generated with seven focal color categories (five chromatic colors: red, yellow, green, blue, purple; and black and white) that were pre-defined as anchor points by an American English speaker in the early 20th century (Munsell, 1907). Constrained by available pigments, Albert Munsell set himself the goal of generating equal supra-threshold perceptual steps between these anchor points. The task was subjective, and although it is replicable, there is still no

objective criterion. The implicit assumption made by Munsell and his followers was that the anchor points are universal, perhaps even hard wired in the nervous system (see (Boynton & Olson, 1990)). But there is no evidence that these anchor points are hardwired by the nervous system. Moreover, there are many so-called perceptually uniform color spaces, each with its own geometry, including spaces produced by the Optical Society of America (Boynton & Olson, 1990), the Natural Color System (Shamey et al., 2011), and the International Commission on Illumination (CIELAB and CIECAM02, (Luo, Cui, & Li, 2006)). The focal-color anchor points chosen by Munsell are salient in Munsell's color space but not especially salient in CIELAB; and the just-noticeable differences are accurate in CIE but less accurate in Munsell. If color space had a universal perceptual origin, how can there be many variations? The alternative account—that color spaces reflect the nature of the tasks used to define them—seems more likely than the notion that there is such a thing as a single, universal, uniform perceptual color space.

The communication-efficiency analysis offers a way out of the conundrum: it provides a more appropriate assay of the underlying forces that determine variability in color naming and color-term evolution because it is not influenced by saturation (Fig. 4). The communication-efficiency framework suggests that the usefulness of color, as determined by task demands, will influence how we represent color space, which presumably will shape how the space is categorized when assessed with a given color space. Thus the communication framework could explain not only the variability in color naming among languages but also the variability in color spaces. And the interaction between perceptual and cognitive mechanisms that is necessarily invoked in creating any color space may explain why establishing a consensus uniform color space has been elusive. The upshot is that any theory of color categories that rests solely on an analysis of a perceptual color space will be incomplete (Zaslavsky et al., 2019).

These considerations redirect the investigation for the causes of color-term variability and evolution toward cognitive factors, which we argue reflect the usefulness of color. What factors account for the patterns in communication efficiency of color among languages? Warm colors might be communicated more efficiently than cool colors across all languages because of the color statistics of natural objects: the things that are identified as objects are more likely to have warm colors (reds, oranges) rather than cool colors (blues, greens) (Gibson et al., 2017; Rosenthal et al., 2018). In particular, the reddish color of faces, attributed to oxygenated blood, is especially important for signaling health, emotion, social status, and sex (Hasantash, Lafer-Sousa, Afraz, & Conway, 2019; Lefevre, Ewbank, Calder, von dem Hagen, & Perrett, 2013; Stephen, Oldham, Perrett, & Barton, 2012). Thus warm colors are likely to be more behaviorally relevant than cool colors across all cultures, which would drive the universal pattern by which warm colors are communicated more efficiently (Gibson et al., 2017). The higher behavioral relevance of warm colors may also explain why there is a relatively larger number of color terms for warm colors than cool colors, as documented in many languages, and may explain why the newest basic color terms such as *peach* are within the warm part of color space (MacLaury, Almási, & Kövecses, 1997; Mylonas, MacDonald, & Wuergler, 2010; Paggetti, Menegaz, & Paramei, 2016; Paramei, Griber, & Mylonas, 2018; Safuanova & Korzh, 2007).

Are yellow/brown and purple, the colors that undergo the most dramatic shifts in communication efficiency, also distinguished by their usefulness? Brown and yellow have the same hue but different brightness (Buck, 2015). As Buck points out, yellow and brown are especially important colors, signaling important features of the environment (earth, rocks, many fruits, feces), as well as more stable features of facial identity such as race (Hasantash et al., 2019; Stephen, Law Smith, Stirrat, & Perrett, 2009)). The imperative to communicate yellow and brown would seem to be substantial, although perhaps not as great as the imperative to communicate warm colors such as those associated with social signaling (health, emotion, sex, social status). Purple,

meanwhile, is a boundary color, situated between warm and cool (Hayter, 1826; Kuehni & Schwartz, 2008). It is extraspectral, relatively rare among stable natural pigments, and prized by royalty. But because of its rarity, it may not have been that useful for much of human history. We hypothesize that technological innovations enabling the production of new pigments have had greater impact on the utility of purple compared to other colors. Future work is needed to test this hypothesis, but anecdotal evidence exists. For example, the first synthetic pigment was purple, produced commercially in 1857 (Travis, 1990). The use of purple in paintings increased substantially in the following years, as evident by the use of purple in impressionist painting. Indeed “violettomania” became a sign of modernism (Reutersvärd, 1950).

The communication-efficiency framework helps reconcile an apparent paradox. On the one hand, the vast majority of language groups appear to carve up a given perceptual color space similarly given a set number of categories and sufficient color-naming data across each population (Gibson et al., 2017; Lindsey et al., 2015; Regier et al., 2007). On the other hand, languages appear to evolve in terms of their number of color categories. How can a language have both a complete representation of color, and be evolving? All people with normal retinal chromatic mechanisms have essentially the same perceptual and discrimination abilities. Indeed, non-human primates have very similar color discrimination mechanisms to humans (Gagin et al., 2014; Horwitz, 2015; Stoughton, Lafer-Sousa, Gagin, & Conway, 2012)—clearly colors can be seen in the absence of language. When presented with a given perceptual space (say Munsell), human observers with normal retinal color mechanisms, regardless of the language they speak, will see the colors similarly, and they will categorize the colors similarly given a specified number of terms (Lindsey & Brown, 2006), even if all people in a population do not all have the words to do so (gathering data from many people in the population and assessing the variability uncovers the universal categorization structure (Gibson et al., 2017; Lindsey et al., 2015)). What varies across cultures and language groups is the extent to which a given color is useful, which accounts for differences in color-categorization systems across the world. As color uses change (or the color spaces used to assess color behavior change), so do color-naming systems. Moreover, as colors become useful, color-naming systems evolve new categories. In this framework, the assertion by Berlin and Kay that there are 11 basic color categories in English is a hypothesis: are these 11 color categories the most useful?

The important role of usefulness in driving color categories is supported by anthropological work showing that color terms are connected to the appearance of specific behaviorally relevant objects. A striking example is provided by the Yéli Dnye term for *black* (Levinson, 2000). The term derives from the name of a tree that has brown nuts that become black when roasted. Thus the color term is linked not just to a specific object, but with the color-defined state of the object in which it is most relevant. This anthropological work shows not only why color categorization depends critically on usefulness but also why the exclusion of specific object labels from the WCS instructions was problematic: color terms invariably derive from objects for which the color is a diagnostic property. The communication-efficiency framework is consistent with observations that some languages simply do not label some color chips (Lindsey et al., 2015), and it predicts that color-naming systems should continue to evolve, to the extent that the relevance of colors changes. This prediction is supported by the evidence described above, suggesting that *cream* and *peach* are emerging as basic color categories in English (Lindsey & Brown, 2014)—these new categories are also, not coincidentally, the names of specific behaviorally relevant objects.

So what can we make of the correlation of saturation and focal-color status? The correlation between focal-color status and saturation evident in the Munsell space may have emerged because Munsell pre-defined the focal-color anchor points. In doing so, Munsell might have

been biased to select more salient samples for these hues, since making explicit category assignments can amplify perceptual biases (Bae et al., 2015). In other words, for Munsell the most useful colors in the space were the ones he defined as the anchor points, so he picked relatively saturated samples for them. Observers of Munsell’s chips would be expected to be sensitive to the saturation of these chips (Helmholtz, 1867 (1909 Edition)Helmholtz, 1867 (1909 Edition); Wyszecki & Stiles, 1982), and find them especially useful because perceptual salience promotes detectability. This account predicts that human infants (lacking mature language) should categorize the Munsell colors according to the anchor points defined by Munsell, which they do (Skellton, Catchpole, Abbott, Bosten, & Franklin, 2017). In our account, the human infant data are not evidence of a biological constraint on categorization, but rather evidence that the human visual system is sensitive to salience. The biological constraint does not bring about the categories observed; instead, the biological constraint governs detectability. Thus although perceptual salience is correlated with color categorization patterns in the Munsell space, the causal link is not from perceptual salience to color naming, but rather by way of the correlation between saturation and usefulness invoked in the construction of the space. Note that we are not arguing that Munsell’s anchor points were arbitrary: presumably he selected them because they correspond to useful categories in early 20th century America (and these categories are likely useful to many cultures around the world). Saturation is, of course, not the only predictor of usefulness, which is how we can account for the emergence of desaturated color categories such as *pink*, *cream*, and *peach*. Desaturated colors are, nonetheless, less salient, which is perhaps why pink has not achieved especially high communication efficiency in any language despite being an accepted basic color term (Fig. 7).

The subjective experience of the importance of the basic color categories is compelling, which may be why explanations for color categories have been sought in the neurophysiology of the visual system. There are three classes of cone photoreceptor in the retina of normal trichromat observers; the cones are called L, M, and S, to indicate the region within the visible spectrum in which each cone type has its peak (L = long wavelength peak; M = middle; S = short). The cone types have historically been referred to as red, green, and blue (so-called primary colors), yet this nomenclature is deeply misleading (Conway, 2009): the cone spectral tuning does not align with these colors. Instead, the spectral tuning of the cones in mammalian trichromats likely reflects selective pressures imposed by the optical constraints of a simple eye, which likely explains why L and M cones have peaks that are only ~30 nm apart and both in the middle, or yellow, part of the visible spectrum (Zaidi, Marshall, Thoen, & Conway, 2014). The compelling nature of primary colors fostered a misunderstanding about how color is encoded in the retina. The primary nature of red, green, and blue (or any set of so-called primary colors) is the computational output of the color-vision system, not a property of the inputs to the system.

An influential theory holds that color depends not on three primary colors but rather on three sets of opponent colors: red-green; blue-yellow; and black-white (Hering, 1964). This theory has often been invoked in support of Berlin and Kay’s framework, yet there is no neural or behavioral data supporting the theory. Three pieces of data have dealt a blow to the validity of Hering’s theory. First, initial microelectrode recordings of responses of neurons in the lateral geniculate nucleus (LGN), corresponding to the first post-receptoral stage of visual processing, were interpreted at one time as the brain basis for the four unique chromatic hues (red, green, blue, yellow) (De Valois, Smith, Kitai, & Karoly, 1958). But color responses of LGN cells do not correspond to these color categories (Webster, Miyahara, Malkoc, & Raker, 2000); instead the LGN color tuning arises because of cone opponency, an essential operation needed to obtain color vision given the high correlation of signals among cone types. Second, there is substantial variability in what colors in a given color space constitute the best exemplars of red, green, blue, and yellow. And third, the primary

psychological evidence in favor of the importance of the unique hues depends on color-cancellation experiments (Hurvich, 1981). Those experiments stipulated the use of the unique hues in the experimental design. It is now clear that binary hues can be used just as successfully in hue cancellation, undermining any claim that the unique hues are privileged (Bosten & Lawrance-Owen, 2014) (Wool et al., 2015). The conclusion of the hue cancellation experiments is that color depends on opponency; the experiments do not stipulate which categories of colors are special.

There may be some subtle neural signatures of the unique hues in visual cortex (Stoughton & Conway, 2008) and behavior (Forder, Bosten, He, & Franklin, 2017), but the weight of cumulative evidence suggests that basic color categories such as the unique hues are not privileged, in either neurophysiology or perception (Bohon et al., 2016; Wool et al., 2015). Despite intensive searches, no brain area has been found that houses a neural population with color tuning properties that correspond to basic color categories. Instead brain areas that likely encode color show a complete representation of color with a slight bias towards warm colors (Bohon et al., 2016; Brouwer & Heeger, 2009; Conway & Stoughton, 2009). We think of this representation as pluripotent, able to support a range of possible demands placed by behavior on our color system. In this conception, basic color categories do not emerge as a result of constraints hard-wired in the brain; they arise because of how color signals are used (i.e. decoded), for example by decision-making centers in the frontal cortex (Bird, Berens, Horner, & Franklin, 2014; Brouwer & Heeger, 2013). The pursuit of psychological explanations that are predicated on the hypothesis that Hering's unique hues are hard-wired in the nervous system seem destined to fail. In sum, the universal aspects of color-naming behavior should prompt a search not for a universal neurobiological or linguistic origin of basic color categories, but rather for a deeper understanding of the role of color in behavior (what is color for?) and the brain mechanisms that render color so useful.

Declaration of Competing Interest

All authors declare that they have no competing interests.

Acknowledgements

Regarding the collection of data, we thank Ricardo Godoy and Tomas Huanca for logistical help; Dino Nate Añez and Salomon Hiza Nate for help translating and running the task; Rashida Khudiyeva and Isabel Rayas for helping with the timing of the reaction times for the color and object naming. For feedback on the work, we thank Evelina Fedorenko, Alexander Rehding, Steven Piantadosi, Roger Levy, and Daniel Garside. The work was supported by NSF IOS Program Award 1353571 (to B.R.C.), Linguistics Program Award 1534318 (to E.G.), and the Intramural Research Program of the NIH (National Eye Institute and National Institute of Mental Health).

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:<https://doi.org/10.1016/j.cognition.2019.104086>.

References

- Abbott, J. T., Griffiths, T. L., & Regier, T. (2016). Focal colors across languages are representative members of color categories. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(40), 11178–11183.
- Baddeley, R., & Atwell, D. (2009). The relationship between language and the environment: Information theory shows why we have only three lightness terms. *Psychological Science*, *20*(9), 1100–1107.
- Bae, G. Y., Olkkonen, M., Allred, S. R., & Flombaum, J. I. (2015). Why some colors appear more memorable than others: A model combining categories and particulars in color working memory. *Journal of Experimental Psychology-General*, *144*(4), 744–763.
- Berlin, B., & Kay, P. (1969). *Basic color terms: Their universality and evolution*. Berkeley, CA: University of California Press.
- Bird, C. M., Berens, S. C., Horner, A. J., & Franklin, A. (2014). Categorical encoding of color in the brain. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(12), 4590–4595.
- Bohon, K. S., Hermann, K. L., Hansen, T., & Conway, B. R. (2016). Representation of perceptual color space in macaque posterior inferior temporal cortex (the V4 complex). *eNeuro*, *3*(4).
- Bosten, J. M., & Lawrance-Owen, A. J. (2014). No difference in variability of unique hue selections and binary hue selections. *Journal of the Optical Society of America A, Optics, Image Science, and Vision*, *31*(4), A357–364.
- Boynton, R. M., & Olson, C. X. (1990). Saliency of chromatic basic color terms confirmed by three measures. *Vision Research*, *30*(9), 1311–1317.
- Brouwer, G. J., & Heeger, D. J. (2009). Decoding and reconstructing color from responses in human visual cortex. *The Journal of Neuroscience*, *29*(44), 13992–14003.
- Brouwer, G. J., & Heeger, D. J. (2013). Categorical clustering of the neural representation of color. *The Journal of Neuroscience*, *33*(39), 15454–15465.
- Buck, S. L. (2015). Brown. *Current Biology*, *25*(13), R536–537.
- Conway, B. R. (2009). Color vision, cones, and color-coding in the cortex. *Neuroscientist*, *15*(3), 274–290.
- Conway, B. R., & Stoughton, C. M. (2009). Response: Towards a neural representation for unique hues. *Current Biology*, *19*(11), 442–443.
- De Valois, R. L., Smith, C. J., Kitai, S. T., & Karoly, A. J. (1958). Response of single cells in monkey lateral geniculate nucleus to monochromatic light. *Science*, *127*, 238–239.
- Forder, L., Bosten, J., He, X., & Franklin, A. (2017). A neural signature of the unique hues. *Scientific Reports*, *7*, 42364.
- Gagin, G., Bohon, K. S., Butensky, A., Gates, M. A., Hu, J. Y., Lafer-Sousa, R., et al. (2014). Color-detection thresholds in rhesus macaque monkeys and humans. *Journal of Vision*, *14*(8), 12.
- Gibson, E., Futrell, R., Jara-Ettinger, J., Mahowald, K., Bergen, L., Ratnasingam, S., et al. (2017). Color naming across languages reflects color use. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(40), 10785–10790.
- Harkness, S. (1973). Universal aspects of learning color codes: A study in two cultures. *Ethos*, *1*, 175–200.
- Hasantash, M., Lafer-Sousa, R., Afraz, A., & Conway, B. R. (2019). Paradoxical impact of memory on color appearance of faces. *Nature Communications*, *10*(1), 3010.
- Haynie, H. J., & Bower, C. (2016). Phylogenetic approach to the evolution of color term systems. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(48), 13666–13671.
- Hayter, C. (1826). *A new practical treatise on the three primitive colours*. London: Booth.
- Helmholtz, H. (1867). In J. P. Southall (Ed.). *Hanbuch der physiologischen optik* (pp. 2). (1909 edition). New York: Dover Publications.
- Hering, E. (1905). In L. Hurvich, & D. Jameson (Eds.). *Outline of a theory of the light sense* Harvard University Press 1964.
- Hering, E. (1964). In L. M. Hurvich, & D. Jameson (Eds.). *Outlines of a theory of the light sense*. Cambridge Massachusetts: Harvard University Press.
- Horowitz, G. D. (2015). What studies of macaque monkeys have told us about human color vision. *Neuroscience*, *296*, 110–115.
- Hurvich, L. M. (1981). *Color vision*. Sinauer Associates Inc.
- Jameson, K., & D'Andrade, R. G. (1997). It's not really red, green, yellow, blue: An inquiry into perceptual color space. In Hardin, & Maffi (Eds.). *Color categories in thought and language* (pp. 295–319). Cambridge, UK: University of Cambridge Press.
- Kay, P., & Maffi, L. (1999). Color appearance and the emergence and evolution of basic color lexicons. *American Anthropologist*, *101*, 743–760.
- Kay, P., & Regier, T. (2003). Resolving the question of color naming universals. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(15), 9085–9089.
- Kohlrash, V. A. (1935). Zur Photometrie farbiger Lichter. *Das Licht*, *5*, 259–275.
- Kuehni, R., & Schwartz, A. (2008). *Color ordered: A survey of color systems from antiquity to the present*. New York, NY: Oxford University press.
- Lantz, D., & Steffire, V. (1964). Language and cognition revisited. *Journal of Abnormal Psychology*, *69*, 472–481.
- Lefevre, C. E., Ewbank, M. P., Calder, A. J., von dem Hagen, E., & Perrett, D. I. (2013). It is all in the face: Carotenoid skin coloration loses attractiveness outside the face. *Biology Letters*, *9*(6), 20130633.
- Leonard, W. R., Reyes-Garcia, V., Tanner, S., Rosinger, A., Schultz, A., Vadez, V., et al. (2015). The Tsimane' Amazonian Panel Study (TAPS): Nine years (2002–2010) of annual data available to the public. *Economics and Human Biology*, *19*, 51–61.
- Levinson, S. C. (2000). Yéi Dnye and the theory of basic color terms. *Journal of Linguistic Anthropology*, *10*(1), 3–55.
- Lillo, J., Gonzalez-Perilli, F., Prado-Leon, L., Melnikova, A., Alvaro, L., Collado, J. A., et al. (2018). Basic color terms (BCTs) and categories (BCCs) in three dialects of the Spanish language: Interaction between cultural and universal factors. *Frontiers in Psychology*, *9*, 761.
- Lindsey, D. T., & Brown, A. M. (2004). Sunlight and "Blue": The prevalence of poor lexical color discrimination within the "grue" range. *Psychological Science*, *15*(4), 291–294.
- Lindsey, D. T., & Brown, A. M. (2006). Universality of color names. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(44), 16608–16613.
- Lindsey, D. T., & Brown, A. M. (2014). The color lexicon of American English. *Journal of Vision*, *14*(2).
- Lindsey, D. T., Brown, A. M., Brainard, D. H., & Apicella, C. L. (2015). Hunter-gatherer color naming provides new insight into the evolution of color terms. *Current Biology*, *25*(18), 2441–2446.
- Lucy, J. A., & Shweder, R. A. (1979). Whorf and his critics – Linguistic and non-linguistic influences on color memory. *American Anthropologist*, *81*(3), 581–615.

- Luo, M. R., Cui, G., & Li, C. (2006). Uniform colour spaces based on CIECAM02 colour appearance model. *Color Research and Application*, 31(4), 304–313.
- MacLaury, R. E. (2007). Categories of desaturated-complex color, sensorial, perceptual, and cognitive models. In MacLaury, Paramei, & Dedrick (Eds.). *Anthropology of color: Interdisciplinary multilevel modeling* (pp. 125–150). John Benjamins Publishing Company.
- MacLaury, R. E., Almási, J., & Kövecses, Z. (1997). Hungarian Piros and Vörös: Color from points of view. *Semiotica*, 114, 67–81.
- Munsell, A. H. (1907). *A Color Notation, a measured color system, based on the three qualities Hue, Value, and Chroma* (2nd edition). Boston: Geo. H. Ellis Co.
- Mylonas, D., MacDonald, L., & Wuerger, S. (2010). Towards an online color naming model. *Color science and engineering systems, technologies, and applications: Eighteenth color and imaging conference*, 140–144.
- Neitz, M., & Neitz, J. (2001). A new mass screening test for color-vision deficiencies in children. *Color Research and Application*, 26, S239–S249.
- Olkkonen, M., Witzel, C., Hansen, T., & Gegenfurtner, K. R. (2010). Categorical color constancy for real surfaces. *Journal of Vision*, 10(9).
- Paggetti, G., Menegaz, G., & Paramei, G. V. (2016). Color naming in Italian language. *Color Research and Application*, 41(4), 402–415.
- Paramei, G. V., D'Orsi, M., & Menegaz, G. (2014). 'Italian blues': A challenge to the universal inventory of basic colour terms. *Journal of the International Colour Association*, 13, 27–35.
- Paramei, G. V., Griber, Y. A., & Mylonas, D. (2018). An online color naming experiment in Russian using Munsell color samples. *Color Research and Application*, 43(3), 358–374.
- Regier, T., & Kay, P. (2004). Color naming and sunlight: Commentary on Lindsey and Brown (2002). *Psychological Science*, 15(4), 289–290 discussion 291–284.
- Regier, T., Kay, P., & Cook, R. S. (2005). Focal colors are universal after all. *Proceedings of the National Academy of Sciences of the United States of America*, 102(23), 8386–8391.
- Regier, T., Kay, P., & Khetarpal, N. (2007). Color naming reflects optimal partitions of color space. *Proceedings of the National Academy of Sciences of the United States of America*, 104(4), 1436–1441.
- Regier, T., Kemp, C., & Kay, P. (2015). Word meanings across languages support efficient communication. In B. MacWhinney, & W. O'Grady (Eds.). *The handbook of language emergence* (pp. 237–263). Hoboken, NJ: Wiley-Blackwell.
- Reutersvärd, O. (1950). The "Violetomania" of the Impressionists. *Journal of Aesthetics and Art Criticism*, 9(2), 106–110.
- Rosenthal, I., Ratnasingam, S., Haile, T., Eastman, S., Fuller-Deets, J., & Conway, B. R. (2018). Color statistics of objects, and color tuning of object cortex in macaque monkey. *Journal of Vision*, 18(11), 1.
- Safuanova, O. V., & Korzh, N. N. (2007). Russian color names Mapping into a perceptual color space. *Anthropology of Color: Interdisciplinary Multilevel Modeling*, 55–74.
- Saunders, B. A. C., & van Brakel, J. (1997). Are there nontrivial constraints on colour categorization? *The Behavioral and Brain Sciences*, 20(2), 167-+.
- Shamey, R., Shepherd, S., Abed, M., Chargualaf, M., Garner, N., Dippel, N., et al. (2011). How well are color components of samples of the Natural Color System estimated? *Journal of the Optical Society of America A, Optics, Image Science, and Vision*, 28(10), 1962–1969.
- Shamey, R., Zubair, M., & Cheema, H. (2019). Unique hue stimulus selection using Munsell color chips under different chroma levels and illumination conditions. *Journal of the Optical Society of America a-Optics Image Science and Vision*, 36(6), 983–993.
- Skelton, A. E., Catchpole, G., Abbott, J. T., Bosten, J. M., & Franklin, A. (2017). Biological origins of color categorization. *Proceedings of the National Academy of Sciences of the United States of America*, 114(21), 5545–5550.
- Steels, L., & Belpaeme, T. (2005). Coordinating perceptually grounded categories through language: A case study for colour. *The Behavioral and Brain Sciences*, 28(4), 469–489 discussion 489–529.
- Steffire, V., Vales, V. C., & Morley, L. (1966). Language and cognition in yucatan – A cross-cultural replication. *Journal of Personality and Social Psychology*, 4(1), 112–8.
- Stephen, I. D., Law Smith, M. J., Stirrat, M. R., & Perrett, D. I. (2009). Facial skin coloration affects perceived health of human faces. *International Journal of Primatology*, 30(6), 845–857.
- Stephen, I. D., Oldham, F. H., Perrett, D. I., & Barton, R. A. (2012). Redness enhances perceived aggression, dominance and attractiveness in men's faces. *Evolutionary Psychology*, 10(3), 562–572.
- Stoughton, C. M., & Conway, B. R. (2008). Neural basis for unique hues. *Current Biology*, 18(16), R698–R699.
- Stoughton, C. M., Lafer-Sousa, R., Gagin, G., & Conway, B. R. (2012). Psychophysical chromatic mechanisms in macaque monkey. *The Journal of Neuroscience*, 32(43), 15216–15226.
- Sturges, J., & Whitfield, T. W. A. (1995). Locating basic colors in the munsell space. *Color Research and Application*, 20(6), 364–376.
- Travis, A. S. (1990). Perkin's mauve: Ancestor of the organic chemical industry. *Technology and Culture*, 31(1), 51–82.
- Tribus, M. (1961). Information theory as the basis for thermostatics and thermodynamics. *Journal of Applied Mechanics*, 28(1), 1–8.
- Webster, M. A., Miyahara, E., Malkoc, G., & Raker, V. E. (2000). Variations in normal color vision. II. Unique hues. *Journal of the Optical Society of America A, Optics, Image Science, and Vision*, 17(9), 1545–1555.
- Witzel, C., Cinotti, F., & O'Regan, J. K. (2015). What determines the relationship between color naming, unique hues, and sensory singularities: Illuminations, surfaces, or photoreceptors? *Journal of Vision*, 15(8), 19.
- Witzel, C., & Gegenfurtner, K. R. (2018). Color perception: Objects, constancy, and categories. *Annual Review of Vision Science*, 4(4), 475–499.
- Wool, L. E., Komban, S. J., Kremkow, J., Jansen, M., Li, X., Alonso, J. M., et al. (2015). Salience of unique hues and implications for color theory. *Journal of Vision*, 15(2).
- Wyszecki, G., & Stiles, W. S. (1982). *Color Science: Concepts and methods, quantitative data and formulae*. New York: John Wiley and Sons, Inc.
- Zaidi, Q., Marshall, J., Thoen, H., & Conway, B. R. (2014). Evolution of neural computations: Mantis shrimp and human colour vision. *Iperception*, 5, 492–496.
- Zaslavsky, N., Kemp, C., Regier, T., & Tishby, N. (2018). Efficient compression in color naming and its evolution. *Proceedings of the National Academy of Sciences of the United States of America*, 115(31), 7937–7942.
- Zaslavsky, N., Kemp, C., Tishby, N., & Regier, T. (2019). Color naming reflects both perceptual structure and communicative need. *Topics in Cognitive Science*, 11(1), 207–219.