



# **ARTICLE**

https://doi.org/10.1057/s41599-019-0341-7

OPFN

# Differences in color categorization manifested by males and females: a quantitative World Color Survey study

Nicole A. Fider 1 \* & Natalia L. Komarova \* \*

ABSTRACT Gender-related differences in human color preferences, color perception, and color lexicon have been reported in the literature over several decades. This work focuses on the way the two genders categorize color stimuli. Using the cross-cultural data from the World Color Survey (WCS) and rigorous mathematical methodology, a function is constructed, which measures the differences in color categorization systems manifested by men and women. A significant number of cases are identified, where men and women exhibit markedly disparate behavior. Interestingly, of the regions in the Munsell color array, the green-blue ("grue") region appears to be associated with the largest group of categorization differences, with females revealing a more differentiated color categorization pattern compared to males. More precisely, in those cases, females tend to use separate green and/or blue categories, while males predominantly use the grue category. In general, the cases singled out by our method warrant a closer study, as they may indicate a transitional categorization scheme.

<sup>&</sup>lt;sup>1</sup> Mathematics Department, University of California, Irvine, Irvine, CA 92697-3875, USA. \*email: nfider@uci.edu; komarova@uci.edu

#### Introduction

t has been asserted (and subsequently, studied for the past several decades) that languages have specific categorization schemes shared among population members, which promote efficient learning and communication among speakers (Berlin and Kay, 1991; Kay and Maffi, 1999; Lindsey and Brown, 2006). Variations in color naming among subgroups of populations have also been studied; see for example, the study of motifs within individual languages and across languages (Lindsey and Brown, 2009). Evolution of color categories in different cultures has been the topic of much discussion (Berlin and Kay, 1991; Dedrick, 1996; Saunders, 2000; Regier et al., 2005). Here we focus on categorization behavior of males and females. Is it possible that these subgroups in the population exhibit significant and systematic differences in their color categorization?

Many studies in the past have demonstrated that there are measurable differences in the way males and females see, perceive, and talk about color. In his seminal work in 1965, Chapanis (1965) found that women were significantly more consistent in matching color chips to color names. In children, there is a minimum age for correct and consistent color-naming, and acquisition among girls is generally faster than among boys (Anyan and Quillian, 1971; Bornstein, 1985). It has further been reported that females have a larger word repertoire and use more elaborate terms to describe color (Simpson and Tarrant, 1991; Nowaczyk, 1982; Greene and Gynther, 1995; Yang, 2001; Mylonas et al., 2014; Lindsey and Brown, 2014). Studies of other languages revealed females commanding a richer color vocabulary than males, including Nepali (Thomas et al., 1978), Chinese (Moore et al., 2002), Caucasus languages (Samarina, 2007), Spanish (MacDonald and Mylonas, 2014), Estonian, Italian, and Turkish (Uusküla and Bimler, 2016), and Russian (Paramei et al., 2018). It has also been documented that females tend to be better than males at matching colors from memory (Pérez-Carpinell et al., 1998) and at retrieving color labels in a speed color-naming task (DuBois, 1939; Saucier et al., 2002; Shen, 2005). In a triad study of Bimler et al. (2004), it was found that males placed less weight on inter-stimulus separation along a red-green axis but more on a lightness axis as compared to females. In a recent web-based psycholinguistic experiment (Griber et al., 2017), in an unconstrained color-naming task, women were found to have a much richer repertory of color words, including a great variety of monolexemic non-basic color terms and "fancy" color names; while men used more basic terms and their compounds. Further it was found that, compared to males, females revealed a more refined linguistic segmentation of color space, predominantly along the red-green axis of color space. Those findings may reflect gender differences in cultural factors relating to range of available color terms and access to them.

A large number of studies have been carried out to determine what types of physiological/perceptual differences might exist between males and females in different aspects of color vision. Differences in unique hue appearance have been reported (Volbrecht et al., 1997; Kuehni, 2001), as well as different performance in color-matching experiments (Birch et al., 1991; Pardo et al., 2007; Haddad et al., 2009), with females exhibiting a larger matching range (Rodríguez-Carmona et al., 2008). In several other recent studies, significant differences in male and female visual functions were reported, related to color appearance and peripheral vision (Abramov et al., 2012). Szeszel et al. (2005) studied the mapping of color words and color appearances among different observer subgroups (defined by perceptual phenotype and photopigment opsin genotype analyses), and found evidence for different representations of linguistic and perceptual similarity across the different groups.

The studies reported above encompass a range of fields, from physiology to psychology and linguistics. It is becoming evident that there are differences in male vs. female processing of color information, that these differences are observed at different levels (from perception to color vocabulary), and can be attributed to a range of mechanisms (from genetic/physiological to behavioral/ social). In the current study, we ask the question of whether these differences may play a role in the evolution of color categorization and emergence of new color categories in individual languages. Therefore, we seek evidence of male-female differences in color categorization across the world cultures. Using the data from the World Color Survey (WCS) (Kay et al., 2009), and the methods previously presented in our work (Fider et al., 2017), we apply rigorous quantitative analysis of color categories by male and female subpopulations, and observe statistically significant differences in male and female responses.

## Methods of analysis

The World Color Survey database has color-naming and focusnaming data from 110 different languages, with an average of 24 tested speakers per language; for a specific language we call the set of language-speakers P and the set of color words elicited by speakers W. To ensure a reasonably sized sample of both males and females, we study only the 91 languages which have at least 8 for each male and female speakers represented in the database.

In our previous work (Fider et al., 2017), we defined a category strength function, CS, which had range [0, 1], and measured the degree of agreement of the population with regards to different color words. We then used this function to identify the set of basic color terms (BCTs) with respect to a threshold value  $t_*$ , which we denote  $W_*$ . For these words, the category strength exceeds a given threshold,  $t_*$ : for  $w \in W_*$ ,  $CS(w) \ge t_*$ .

We can further identify two cut-off values,  $t_{\rm low}$  and  $t_{\rm high}$ , which allow one to judge just how strong the CS function is for different color terms. If  $CS(w) < t_{\rm low}$ , we say that the corresponding color term is *never-basic*. Further, if  $CS(w) > t_{\rm high}$ , then the term is *always-basic*. Finally, for the intermediate cases with  $t_{\rm low} < CS(w) < t_{\rm high}$ , we say that the term is *potentially-basic*. Based on the statistics of the WCS, we identified  $t_{\rm low} = 0.168$  and  $t_{\rm high} = 0.334$  (see Fider et al., 2017). For convenience, we will sometimes use the non-italicized "basic" to describe *always-basic* or *potentially-basic* color terms, and "nonbasic" to describe *never-basic* color terms.

Now we apply the methods and ideas of Fider et al. (2017) to the male and female subpopulations separately. Let  $P_F$  denote the set of female speakers in the population. For all  $w \in W$ , we define

$$r_w(i) = \{ \text{all chips called w by } i \in P_F \}.$$

Then for two female observers, i,  $j \in P_F$ ,  $i \neq j$ ,

$$r_w(i) \cap r_w(j)$$

gives the set of color chips called w by both female speakers i and j. We can summarize pairwise agreement on word w across the entire female population with

$$\text{CS}_{\text{F}}(w) = \frac{1}{2|P_F|(|P_F|-1)} \sum_{i,j \in P_F, i \neq j} \left( \frac{|r_w(i) \cap r_w(j)|}{|r_w(i)|} + \frac{|r_w(i) \cap r_w(j)|}{|r_w(j)|} \right)$$

By construction,  $CS_F(w) \in [0, 1]$  for all  $w \in W$ . This function measures the female population agreement as to how word w is used. Using function  $CS_F$  and the threshold values from Fider et al. (2017), we can identify the color words, and therefore the corresponding color categories, which are *always-*, *potentially-*, or *never-basic* with respect to the female subpopulation's colornaming behavior.

The construction of the category-strength function for the male subpopulation of a language,  $CS_M$ , is similar. This function measures the male population agreement as to how word w is used. We can use  $CS_M$  to identify color words, and therefore the corresponding color categories, which are *always*-, *potentially*-, or *never-basic* with respect to the male subpopulation's colornaming behavior.

Comparing the male/female color categories. For any word w in a given language, one can define an individual's term map based on his or her usage of the word w in the color-naming task. One can also compile a population term map for w based on the aggregate usage of the word w, as done in Fider et al. (2017) and Kay et al. (2009). The term map of w according to a population can be represented as function  $TM_w$ :  $C \rightarrow [0, 1]$ , with

$$c \mapsto \frac{|\{p \in P | p \text{ called } c \text{ by } w\}|}{|P|}$$
 (1)

where P is the set of population members and C is the set of colors (or colored chips). The term map is defined for a given color word, and assigns for each chip the fraction of the population that uses the word for this chip.  $TM_w$  can be visualized using a two-dimensional heatmap when the underlying color space is chosen to be two-dimensional.

By restricting our study to only the female/male subpopulations, we can compile the population term maps according to female/male speakers of a language. That is, for each  $w \in W$ , we can construct  $TM_{w,F} \colon C \to [0,1]$  and  $TM_{w,M} \colon C \to [0,1]$  based on Eq. (1).

To study how differently males and females of a population use word *w*, we define and use the following difference function:

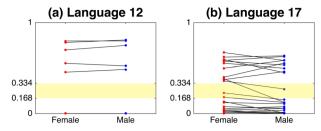
$$\mathrm{Diff}\left(\mathrm{TM}_{\boldsymbol{w},\mathrm{F}'}\mathrm{TM}_{\boldsymbol{w},\mathrm{M}}\right) = \frac{\sum_{\boldsymbol{c} \in \mathrm{supp}(\boldsymbol{w})} \left|\mathrm{TM}_{\boldsymbol{w},\mathrm{F}}(\boldsymbol{c}) - \mathrm{TM}_{\boldsymbol{w},\mathrm{M}}(\boldsymbol{c})\right|}{|\mathrm{supp}(\boldsymbol{w})|} \quad (2)$$

where  $\operatorname{supp}(w)$  is the set of colors  $c \in C$  such that c is relevant to either males and females—that is, if more than one male named c with w or more than one female named c with w, then c is in  $\operatorname{supp}(w)$ . By construction, the largest value Diff  $(\operatorname{TM}_{w,F}, \operatorname{TM}_{w,M})$  can take is 1. The smallest value Diff  $(\operatorname{TM}_{w,F}, \operatorname{TM}_{w,M})$  can take is 0.

#### Results

Category strength and term maps. We used our methodology to identify BCTs for male and female populations separately. Plots of term maps are presented in Supplementary Section 7, Figs S7–S97, for all the languages in WCS with at least eight male and eight female respondents. Considering all color words across all WCS languages, in many cases, the male and female subpopulations do seem to utilize similar word-usage. We are, however, interested in studying cases where male and female naming behaviors appear to be very different. To illustrate some possible similarities and differences in category-strength between males and females, we start with two examples. Figure 1 shows the female and male category strength values, CS<sub>F</sub> and CS<sub>M</sub>, side by side for two WCS languages.

The left panel shows the strengths of the categories with respect to the male and female subpopulations of the Bauzi language (L12 in the WCS archives). Each red (blue) point represents a color word and is plotted at a height corresponding to its female (male) category strength, CS<sub>F</sub> (CS<sub>M</sub>) value. Points which correspond to the same word are connected by black lines. Of the seven color words used by Language 12 speakers, five are classified as *alwaysbasic* color terms (the two *never-basic* color terms have category strength 0), and even when separated into male and female subgroups the five words are *always-basic* with respect to both subpopulations; the ordering of the category strength of the terms



**Fig. 1** Visualization of term strengths. **a** shows data for Language 12, **b** shows data for Language 17. In each subfigure, each color word w of the language is plotted in blue at height  $CS_M(w)$ , and in red at height  $CS_F(w)$ . Lines are drawn to connect male and female results for word w. Points that fall within the yellow region of each graph correspond to *potentially-basic* color words; points above correspond to *always-basic* color words, and points below correspond to *never-basic* color words

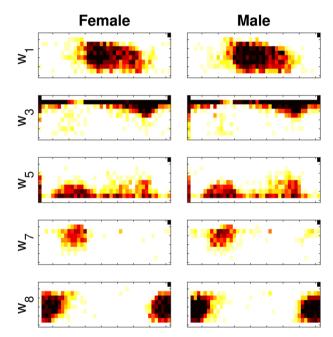
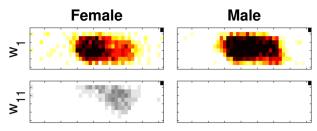


Fig. 2 Language 12's subpopulation term maps for gender-separated data. Each row corresponds to a word, with the left panel showing the female term map, and the right panel showing the male term map

is very similar in males and females. The population term maps for male and female subpopulations, which we denote  $TM_{w,F}$  and  $TM_{w,M}$  for each category corresponding to a word w, are shown as heatmaps in Fig. 2. Each rectangular pixel in the term maps represents a color chip used in the WCS, such that the pixels in the term maps and the WCS grid chips are oriented in the same way; the full WCS chip set is shown in Fig. S1 in Supplementary Section 1 for reference. Darker shading in the term maps indicates that the corresponding colors are named with the word  $w_i$  by a larger fraction of the subpopulation. White coloring in the term maps indicates that the corresponding colors are never named with the word  $w_i$ . We can see that the naming behavior of female and male speakers match closely on all five relevant color terms

The male and female subpopulations show different behaviors in the Cakchiquel language (L17 in the WCS archives). In Fig. 1, right panel, we can see that the orderings of terms according to gender-derived category strengths are very different. We will



**Fig. 3** Real and simulated Diff data. **a** Histogram of Diff value distribution. **b** Histogram showing the number of cases with Diff ≥ 0.25 (non-horizontal axis), for random splitting of the population into pseudofemales and pseudomales (normalized to have a unit area). The tail, highlighted in orange, represents the percentage of iterations which had 19 or more of such "interesting cases"

return to this language in the section when we explore concrete differences in male and female color-naming behavior.

To get a more global view of how generally similar or different male and female naming behaviors can be, we quantify the differences between  $TM_{w,F}$  and  $TM_{w,M}$  using the function Diff  $(TM_{w,F}, TM_{w,M})$  across all languages, and across all color words which are basic with respect to at least one of the genders. Figure 3 shows a histogram of all of these data. We can see that while most color terms show similar male/female naming behavior, there are terms where the difference is relatively big; there are 19 color words, spanning 14 languages, which have Diff  $(TM_{w,F}, TM_{w,M})$  values larger than 0.25.

An important question is whether the large differences in color categorization obtained here are really a signature of differences in female and male behavior. It may be possible that any random splitting of a population in two groups is likely to give some differences in categorization (or, term-map appearance) just by chance. To study the statistical significance of the results of Fig. 3, we randomly divided each population into two subgroups (the pseudomale group and pseudofemale group) and applied the same methodology to find the global distribution of differences between the pseudomale and pseudofemale naming behaviors of each language. We then counted and recorded the number of terms that have Diff value above 0.25. We did this 10,000 times; Fig. 3 shows the the distribution of counts obtained, with area normalized to equal 1 unit.

Now assume the following null hypothesis: "19 or more cases with  $Diff \ge 0.25$  can be obtained by randomly dividing each population into two subgroups". Note that when normalized to have a unit area, the histogram in Fig. 3 can be interpreted as a probability distribution which shows the likelihood of n high-Diff terms appearing. If the null hypothesis is correct, the "tail" of the normalized histogram, highlighted in orange, would have area larger than or equal to 0.05 (using the 95% cut-off). This is however not the case—the area observed is approximately 0.0191, which implies that a large degree of difference in naming behavior, assuming subgroups are formed by random splitting of the general population, occurs with a very small probability. We can therefore conclude that the differences observed by studying males and females are statistically significant.

Analysis of the 19 terms satisfying Diff (w) > 0.25 reveals that three terms come from the Karaja language (L53 in the WCS archives): ikura, iura, and idy. It is noted in Kay et al. (2009) that collecting color-naming data was irregular for this language—data was collected in groups rather than from individuals, which causes the individual, subpopulation, and full-population term maps to exhibit unusual distributions. We therefore omit this language from study; note that omitting L53 from all simulation runs still yields a histogram with tail size less than 0.05, meaning

Table 1 Information on the 13 languages analyzed as case studies, organized by assigned world color survey language number

L (#)	Language name	Country	Language family
6	Ampeeli	Papua New Guinea	Trans-New Guinea
17	Cakchiquel	Guatemala	Macro-Penutian
21	Cavinena	Bolivia	Tacanan
30	Colorado	Ecuador	Macro-Chibchan
34	Djuka	Surinam	(Creole)
45	Huastec	Mexico	Macro-Penutian
46	Huave	Mexico	Macro-Penutian
49	lwam	Papua New Guinea	Sepik-Ramu
67	Mahahua	Mexico	Oto-Manguean
75	Murle	Sudan	Nilo-Saharan
81	Patep	Papua New Guinea	Austronesian
94	Tboli	Philippines	Austronesian
103	Walpiri	Australia	Australian
Data abou	ıt language origin and fami	ly taken from Cook et al. (20	12) and Bimler (2007)

that excluding L53 does not change the conclusions of the statistical significance analysis.

If we had chosen to observe terms with Diff (w) > 0.2, we would have found 79 terms in the WCS data set. This is too many terms to study on a case-by-case basis in this paper, although we highlight some special examples from this set in the Discussion. By performing additional significance analysis, we find that less than 1% of in 10,000 yield 79 or more terms with Diff (w) > 0.2, so we can conclude that the study of this set of terms is also relevant.

Case studies: large differences between female and male term maps. After removing L53 from study, we are left with 16 words across 13 languages; detailed information for these languages is provided in Table 1. Below we explore these 16 terms of interest. Considering the color maps of languages that appear on this list, we can identify the following key groups determined by large male–female differences:

- One gender lexicalizes a category or a category split, and the other gender does not (languages 75, 81, 30, 94, 103, 67). This could be caused by one gender learning/acquiring a category before the other gender, and in our high-Diff cases this happens in the "green/blue/grue" region of the color space. Conventionally, "grue" refers to the collection of colors that can be described in English by either blue or green.
- The genders lexicalize similar categories, but may have different preferred names for the category. This could be caused by native color-word synonyms (language 103), or borrowed color-words which compete with existing native color-words (languages 67, 45, 17). In our high-Diff cases, we see this occurring in the "purple" region and (one example) in the 'green/blue/grue" region of the color space.
- Other (languages 6, 21, 34, 46, 49).

Below we consider in greater detail the cases that indicate the emergence of a category in only one gender's color categorization scheme. Specifically, we highlight the color-naming behavior exhibited by the Murle, Patep, Colorado, Tboli, Walpiri, Mazahua, Huastec, and Cakchiquel languages. The remaining large-Diff cases (termed "others" above) do not exhibit unusual or interesting behavior so we address them in Supplementary Section 2.

Case 1: Murle (L75). The Murle language has one term, nyapus  $(w_{11})$ , which has a high Diff value, see the second row of Fig. 4.

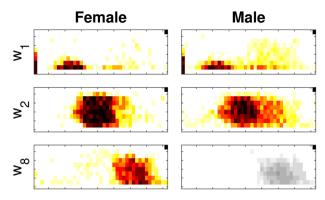


Fig. 4 Selected categories from Language 75. Here and below, gray color maps indicate *never-basicness* of the category

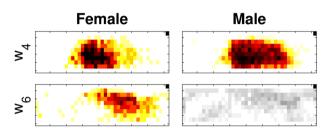


Fig. 5 Selected categories from Language 81

We can see that according to the female subpopulation,  $w_{11}$  occasionally designates the "light blue" region of colors, while the male subpopulation does not use  $w_{11}$  at all. Male speakers use only  $w_1$  to designate grue colors. While females also use  $w_1$  to designate grue colors (the grue category), the terms maps seem to indicate that the female subpopulation used a weak extra category covering the "light blue" colors, which the male subpopulation does not use. We also notice that the grue category for females is polarized toward the green hues, while the male "grue" is relatively balanced.

Case 2: Patep (L81). The Patep language has one term, bilu ( $w_8$ ), which has a high Diff value, see the third row of Fig. 5. We can see that  $w_8$  best designates the "blue" region of colors. However, we can also see by the grayscale coloring of the male  $w_8$  category that male speakers do not use  $w_8$  often enough or consistently enough to qualify  $w_8$  as CS<sub>M</sub> basic. Indeed, male speakers used  $w_2$  to designate "blue" and "green" colors, and also occasionally use  $w_1$  to designate the "green" colors. Females separate the "green" and "blue" color categories distinctly with  $w_2$  and  $w_8$ , respectively.

Case 3: Colorado (L30). The Colorado language has one term, losimban  $(w_4)$ , which has a high Diff value, see the first row of Fig. 6. Females use  $w_6$  to designate "blue" and  $w_4$  to designate "green". In contrast, males rarely use  $w_4$  to designate both "green" and "blue" colors;  $w_6$  is used quite rarely by the male population and seems to be used to designate the colors which do not fall into a known category.

Case 4: Tboli (L94). The Tboli language has one term, gingung  $(w_7)$ , which has a high Diff value (see Fig. 7). Females use  $w_7$  to designate "dark blue-purple" while males rarely use  $w_7$ ; "dark blue-purple" colors are not represented in any other male category.

In the next case studies, we observe the coexistence of competing names for the same color category.

Case 5: Walpiri (L103). The Walpiri language has one term, wajirrkikajirrki ( $w_{12}$ ), which has a high Diff value, see the second

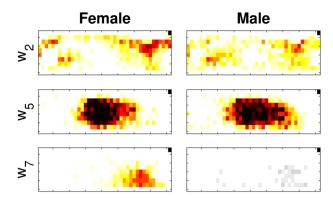
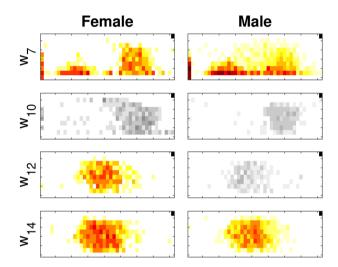


Fig. 6 Selected categories from Language 30



**Fig. 7** Selected category from Language 94. Female and male term maps, respectively, corresponding to color term  $w_7$ 

row of Fig. 8. Females have two competing words which designate "green" colors:  $w_{12}$  and  $w_{14}$ . The "black" and "blue" colors are covered by  $w_7$ . On the other hand, males rarely use  $w_{12}$  to designate "green", but use  $w_{14}$  (and very occasionally  $w_7$ ) to designate "green". Other than a weak presence in the  $w_7$  category, "blue" colors appear in the nonbasic category  $w_{10}$  for both males and females (with a higher strength for females). Walpiri was considered in some detail in Lindsey and Brown (2009) who identified the existence of five color-naming motifs in this language; it appears that gender differences in color naming contribute to this diversity.

Case 6: Mahahua (L67). The Mazahua language has two terms which have high Diff values: morado, and verde. We refer to the words, respectively, as  $w_{28}$  and  $w_{47}$ , based on the WCS enumeration. Male and female term maps for  $w_{47}$  are shown in the sixth row of Fig. 9. We can see that male speakers almost never use  $w_{47}$  to designate any colors, while females use  $w_{47}$  with high frequency and consistency when describing colors which approximate the English "green" category. This is especially interesting when we consider the term maps of  $w_4$ , shown in the second row of Fig. 9. Females use  $w_4$  to designate English "blue" colors, while males use  $w_4$  to designate the combination of "blue" and "green" colors ("grue"). Therefore, this is an example where one gender lexicalizes a large category ("grue"), while the other gender divides it into two smaller categories ("blue" and "green"). Male and female term maps for  $w_{28}$  are shown in the fifth row of Fig. 9.  $w_{28}$  is used by males and females to designate the "purple"

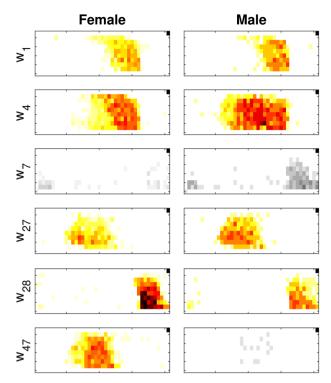


Fig. 8 Selected categories from Language 103

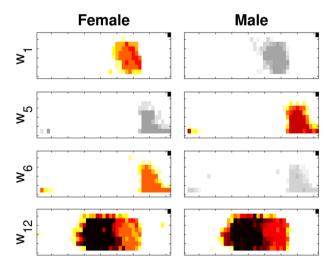


Fig. 9 Selected categories from Language 67

region of colors. However, female speakers use only  $w_{28}$  to designate "purple" while male speakers also use  $w_7$  to designate the same set of colors.

Case 7: Huastec (L45). The Huastec language has two terms, morado and muyaky ( $w_5$  and  $w_6$ ), which have high Diff values (see Fig. 10). Females and males use both terms to designate the "purple" region of colors. However, female speakers favor  $w_6$  while male speakers favor  $w_5$ . It is interesting that in this language the males use the term morado, borrowed from Spanish, whereas females use (traditional) muyaku. This shows a pattern similar to that found by Samarina (2007) in languages of Caucasus, which is explained by gender differences in the life-style. Females who are typically involved in practices requiring attention to foods, dyes, and plants, tend to use indigenous, descriptive color terms. Males, in contrast, get involved in trade and other activities beyond

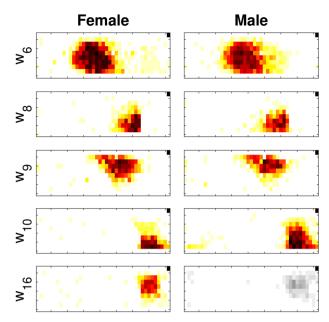


Fig. 10 Selected categories from Language 45

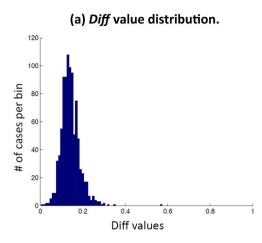
domestic environment, which leads to them using more abstract, adopted color terms.

Case 8: Cakchiquel (L17). The Cakchiquel language has one term, lila ( $w_{16}$ ), which has a high Diff value, see the fifth row of Fig. 11. Females use  $w_{16}$  to designate "light purple"; males use  $w_{16}$  with less frequency and consistency when describing the same set of colors. However, we can see that while females use  $w_{10}$  to designate "dark purple" colors, males use  $w_{10}$  to designate all colors in the "purple" region, including the light and dark varieties.

Male and female category exemplars. In Fider et al. (2017) we outline methods for identifying and analyzing category exemplars according to data from a color-naming task. Applying those methods to the female and male subpopulations, we found that although in some languages, male and female exemplars were different, this result was not statistically significant, in the sense that similar patterns were observed in simulations with randomly selected pseudomale and pseudofemale populations. It should be noted that the algorithm which locates the exemplar of a category depends on finding the three-dimensional centroid of a collection of colors and projecting it back onto the WCS color set. The original collection of colors comes from the WCS color set, which is chosen primarily from the "surface" of a three-dimensional color solid; computing the center of mass and projecting back onto the WCS grid introduces the potential for error, and the seemingly random male/female exemplar results could simply be a consequence of this issue. Details regarding our exemplar-based methods and results can be found in Supplementary Section 4.

#### **Discussion**

Systematic computational analysis of the WCS data revealed the existence of a number of differences in color categorization systems by males and females. The differences that we observed are of several kinds, including cases where (i) one of the genders uses two terms and the other only term one for a given color region, as in Case 2 (L81); (ii) one of the genders has a BCT and the other does not for a given region, as in Case 5 (L103); (iii) the two genders strongly favor different words to describe the same color set, as in Case 7 (L45).



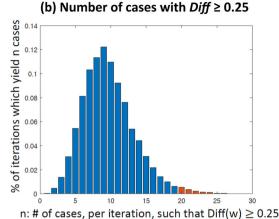


Fig. 11 Selected categories from Language 17

These findings contribute to the relatively large body of literature describing male-female differences in other aspects of color-related behavior. Several reasons for these differences have been put forward, which can be classified as pertaining to "nature" or "nurture" see Mylonas et al., (2014). Genetic differences between males and females have been ascribed to heterozygosity in the X-chromosome genes coding for cone photopigments (Rodríguez-Carmona et al., 2008). The inherited dimorphisms of the genes that encode retinal long- (L) and middle- (M) wavelength photopigments are thought to result in more refined color perception and enhanced ability to discriminate color differences along the red-green axis (Jameson et al., 2001; Jordan et al., 2010; Murray et al., 2012). Szeszel et al. (2005) demonstrated that females with dimorphism of both L-opsin and M-opsin genes exhibited significantly higher consensus in tasks involving judgment of colors or color terms, compared to the other three female genotypes investigated (with no opsin gene diversity or with dimorphism of either L-opsin or M-opsin gene). Further, it has been hypothesized that females have a higher chance to be heterozygous carriers of deutan color vision deficiencies, and thus potentially have a genetic make-up for tetrachromacy (Jameson et al., 2001; Jordan et al., 2010). Gender-related genetic variation in the opponent system responses has also been proposed as a possible cause of male-female differences in performance, such as in the findings of Kuehni (2001), where female observers revealed a larger range of unique hues compared to their male counterparts. The role of testosterone was hypothesized in Abramov et al. (2012), who suggested that it affected the process of recombination and re-weighting of neuronal inputs from the lateral geniculate nucleus to the cortex.

On the other hand, socialization and behavioral patterns have been used to explain some of the observed gender differences. Bimler et al. (2004, p. 128) suggest that the male-female differences in the size of color vocabulary, the fluency in finding color samples to match color terns, and the ability to match colors from memory, could be accounted by the divergent patterns of socialization for males and females that "instill a greater awareness of color among women". Hurlbert and Ling (2007) mention the importance of the evolutionarily different roles of females and males in the society (the gatherers and hunters, respectively), such that females needed better discrimination to detect reddish fruits against a greenish foliage. Greene and Gynther (1995) explain the superior performance of women in color tasks by the differential socialization for women and men, including different clothing, hobbies, and occupations. Yang (1996) reports a direct correlation between subjects' performance on color tasks and their "color" hobbies scores, and mentions that men have significantly

fewer color-related hobbies than women (see also Simpson and Tarrant, 1991). Large gender-related differences in more traditional societies are discussed in Samarina (2007, p. 463), who writes, in the context of Caucasus languages, that "the sources of formation of special female "color subculture" <...> can be found in the division of practice domains between men and women," with women engaging more in activities with access to dyes, foods, indigenous raw materials, plants, etc. Any and all of these factors may contribute to the differences reflected in the WCS categorization data, given the complexity of categorization task, which involves both physiological and psychological layers.

It is intriguing to notice that studies on differences in male-female color perception suggest that the largest variation occurs in the middle of the spectrum, associated with the greenish tones. Abramov et al. (2012) presents a study of basic visual functions, such as color appearance, without reference to any objects. It is determined that males have a broader range of poorer discrimination in the middle of the spectrum (greenish tones), compared to females. Further, in a color-matching study of Murray et al. (2012), females showed substantially less saturation loss than males in the greenish region of color space.

Our results show that of the regions in the Munsell color array, the green-blue region appears to be associated with the largest group of categorization differences. This so-called grue (green and blue) category has received much attention in the literature (see e.g. Lindsey and Brown, 2002; Jameson, 2005; Hardy et al., 2005; Swinkels, 2015). In Lindsey and Brown (2004), the authors analyze the location of the foci of the category grue and find that in some languages the grue focus coincides with the blue focus, while in others it coincides with the green focus. In the third group of languages found by Lindsey and Brown (2004), the grue focus is placed between the green and blue foci, suggesting that the speakers truly do not distinguish green and blue subcategories. Lindsey and Brown (2009) analyzed the WCS data to discover the existence of a small number of "motifs", among which "Dark", "Gray", "Grue", and "Green-Blue-Purple" ("GBP") were the most frequent. All the languages were mapped onto a simplex in accordance with their most frequent motifs. As expected, there were stable languages near the vertices of the simplex, and there were also languages mapping onto the edges connecting the vertices, representing languages "in transition". The transition "Dark"  $\rightarrow$  "Grue"  $\rightarrow$  "GBP" could be seen from this representation, which is consistent with Berlin and Kay's stages of color term evolution (Berlin and Kay, 1991). In addition, there were languages that mapped onto the interior of the simplex; those languages contained multiple motifs and appear to follow a more complex trajectory, see also Kay et al. (1997) and Kay and Maffi (1999).

Diff	Lang	Selected term	Females	Males	M or F
0.34	L67	w <sub>47</sub> (green)	Use this and other term, with blue-centered grue	only use 1 green, grue balanced	F
0.33	L81	w <sub>8</sub> (extended blue)	BCT, more green-biased grue	NBCT, somewhat green-biased grue	F
0.32	L75	w <sub>11</sub> (light blue)	NBCT and green-centered grue	do not have a term, grue bal-anced	F
0.30	L94	w <sub>7</sub> (blue-purple)	BCT and green-centered grue	NBCT and balanced grue	F
0.25	L103	w <sub>12</sub> (green)	BCT + stronger NBCT blue	NBCT + weaker NBCT blue	F
0.24	L58	w <sub>6</sub> (green-centered grue)	NBCT	BCT	M
0.23	L28	w <sub>9</sub> (blue)	BCT	NBCT, and use extended black	F
0.23	L34	w <sub>21</sub> (light blue)	NBCT + another blue	BCT + another blue	M
0.22	L30	w <sub>6</sub> (extended blue)	BCT and green-centered grue	NBCT and balanced grue	F
0.22	L87	w <sub>1</sub> (green)	NBCT and balanced grue	BCT and blue-focused grue	M
0.20	L3	w <sub>12</sub> (green-focused grue)	NBCT	BCT	M

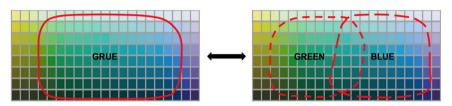


Fig. 12 A schematic illustrating a "grue" category and "blue"/"green" categories. We discuss the possibility of languages transitioning from one scheme to the other

Given our present findings of gender-different categorization of color space in the grue region, we will assume that if a language's categorization scheme is shifting from one which lexicalizes grue to one which lexicalizes green and blue, then the unique grue category can shift its focus toward the green (or blue) region, with a new term simultaneously appearing with the center nearer the blue (or green) hues. This is consistent with the picture that emerges from the analysis of the WCS data. It is important to point out, however, that the WCS data are static, and do not carry temporal information—it is therefore also possible for a categorization scheme to evolve into a less complex one, such as when a color category is discarded.

In Table 2, we list all the languages that have a category with a large difference between female and male categories (the Diff value above 0.2), such that this category is in the grue region. When ordered by the Diff values, the first five such languages exemplify cases, where the females have a more complex color categorization pattern in the grue region. More precisely, females use separate green and blue categories in two and three cases, respectively, while males predominantly use the grue category. In the first four cases, the the focus of grue for the males is relatively balanced, while the focus of female grue is shifted away from a weak green or blue term.

While each given example of one gender having a stronger term than the other is not surprising and can be attributed to chance (and to a relatively small number of informants), the fact that the largest differences correspond to females exhibiting more complexity is consistent with an overall hypothesis that they tend to use finer categories in the grue region.

If one thinks of the WCS data as a snapshot of language evolution, different languages exemplify different stages of color category development. Furthermore, if one accepts the premise that a fundamental purpose of categorization schemes is to allow members of a population to communicate with each other effectively, it makes sense to suppose that differences in individual and subpopulation schemes will ultimately evolve and converge

to a single population scheme (see Fig. 12). The WCS provides a glimpse into such events, by "catching" those languages just at the right moment, when there is an observable difference in subpopulation categorization behavior, and before the overall population scheme is stabilized.

In some cases, as noted in Kay et al. (2009), there is ample evidence regarding the directionality of category evolution for certain languages. In such instances, the origins of a new category's name may be local, or borrowed from other languages. In the latter case, the phenomenon of linguistic acculturation resulting from extended cultural contact and individual bilingualism has been observed, which has been demonstrated to influence color vocabulary and categorization (Hickerson, 1971). For example, Tzeltal has been subject to the influence of Spanish for more than 300 years; traditional term yas corresponds to grue, but because of a borrowed term azul, there is a variation in the usage of yas with a tendency to restrict it to the green region (Berlin and Kay, 1991). Another example comes from Basque, where the traditional grue term became restricted to "blue", with the borrowing of grue ("green") and grue ("gray") from Spanish (cf. verde, gris) (Miller, 2014). It has been noted that specifically the color word for "blue" is often a loan word (Berlin and Kay, 1991). Examples include the borrowing of the English blue by many African languages (sometimes transforming it to bru). The Battas of Sumatra use the word balau borrowed from Dutch. Berbers use samawi (sky color) borrowed from Arabic. In the languages exemplified in the present work, the new terms adopted predominantly by female speakers are borrowed in the case of language L81 (bilu for blue), L67 (morado for purple, verde for green), L45 (azul for blue), L17 (lila for purple), and others; see Kay et al. (2009).

As already mentioned, for the majority of cases in the WCS, there is no information to indicate if and how a language's color categories are shifting. One can however speculate that, based on the notion that females are the vanguard of language development (Labov, 1990; Milroy and Milroy, 1993; Holmes, 1997;

Nevalainen, 2000), it is perhaps the female categorization that will be adapted by the future generations. If this is true, then one could say that in these cases there is a female-driven split of the grue category into separate blue and green categories. This idea echoes an important recent finding of Lindsey and Brown (2014), who focused on color motifs in American English. Two prevalent motifs were found, one the familiar "green-blue" motif of the WCS and the other, a novel "green-teal-blue" motif, which included an extra color (teal) in the grue region, as well as three other high consensus terms (peach, lavender, and maroon). Interestingly, women were significantly more likely to use the "green-teal-blue" motif that contained more color terms. It was hypothesized that "language related to color is changing, and that women are in the vanguard," although the authors cautioned that historical data would be necessary to test this theory.

Using the methods of this paper, we are able to identify specific terms and languages, where the categories exhibited by female and male speakers are very different. These are the cases that warrant a closer study, as they may indicate a transitioning categorization scheme. In the most interesting cases, we find that the two subpopulations can utilize differing categorization patterns, with varying degrees of complexity—for where one gender lexicalizes a single category, for example, the other might lexicalize two, thereby using a more complex category scheme. The most common example found in this paper is in the grue region of colors, where females tend to use the more complex scheme.

#### **Data availability**

Data are available at http://www1.icsi.berkeley.edu/wcs/.

Received: 5 March 2019; Accepted: 28 August 2019;

Published online: 12 November 2019

# References

- Abramov I, Gordon J, Feldman O, Chavarga A (2012) Sex and vision II: color appearance of monochromatic lights. Biol Sex Differ 3(1):21
- Anyan Jr WR, Quillian WW (1971). The naming of primary colors by children. Child Dev 42:1629–1632
- Berlin B, Kay P (1991). Basic color terms: their universality and evolution. University of California Press
- Bimler D (2007) From color naming to a language space: an analysis of data from the World Color Survey. J Cognit Cult 7(3):173–199
- Bimler DL, Kirkland J, Jameson KA (2004) Quantifying variations in personal color spaces: are there sex differences in color vision? Color Res Appl 29 (2):128–134
- Birch J, Young A, David S (1991) Variations in normal trichromatism. In: Drum B, Moreland JD, S. A eds Colour vision deficiencies X. Documenta Ophthalmologica Proceedings Series, vol 54. Springer, Dordrecht, pp 267–272
- Bornstein MH (1985) On the development of color naming in young children: data and theory. Brain Lang 26(1):72–93
- Chapanis A (1965) Color names for color space. Am Sci 53(3):327-346
- Cook R, Kay P, Reiger T (2012) World Color Survey data archives, http://www1. icsi.berkelev.edu/wcs/dataro.html
- $\label{eq:definition} Dedrick\ D\ (1996)\ Color\ language\ universality\ and\ evolution:\ on\ the\ explanation\ for\ basic\ color\ terms.\ Philos\ Psychol\ 9(4):497–524$
- DuBois PH (1939). The sex difference on the color-naming test. Am J Psychol 52:380–382
- Fider N, Narens L, Jameson KA, Komarova NL (2017) Quantitative approach for defining basic color terms and color category best exemplars. J Optical Soc Am A 34(8):1285–1300
- Greene KS, Gynther MD (1995) Blue versus periwinkle: color identification and gender. Percept Mot Skills 80(1):27–32
- Griber YA, Paramei GV, Mylonas D (2017). Gender differences in Russian colour naming. In: Lee Y, Hwang J, Suk HJ, Park YK (eds) Proceedings of the 13th congress of the International Colour Association (AIC 2017): being color with health, Jeju, Korea, pp 0S05-04
- Haddad HJ, Jakstat HA, Arnetzl G, Borbely J, Vichi A, Dumfahrt H, Renault P, Corcodel N, Pohlen B, Marada G et al. (2009) Does gender and experience influence shade matching quality? J Dent 37:e40-e44

- Hardy JL, Frederick CM, Kay P, Werner JS (2005) Color naming, lens aging, and grue: what the optics of the aging eye can teach us about color language. Psychol Sci 16(4):321–327
- Hickerson NP (1971) Review of basic color terms: their universality and evolution by Brent Berlin and Paul Kay. Int J Am Linguist 37(4):257–270
- Holmes J (1997) Setting new standards: sound changes and gender in New Zealand English. Engl World-Wide 18(1):107–142
- Hurlbert AC, Ling Y (2007) Biological components of sex differences in color preference. Curr Biol 17(16):R623–R625
- Jameson KA (2005) Why GRUE? An interpoint-distance model analysis of composite color categories. Cross-Cult Res 39(2):159–204
- Jameson KA, Highnote SM, Wasserman LM (2001) Richer color experience in observers with multiple photopigment opsin genes. Psychon Bull Rev 8 (2):244–261
- Jordan G, Deeb SS, Bosten JM, Mollon J (2010) The dimensionality of color vision in carriers of anomalous trichromacy. J Vis 10(8):12–12. https://doi.org/ 10.1167/10.8.12
- Kay P, Berlin B, Maffi L, Merrifield W, Cook R (2009) The world color survey. CSLI, Stanford
- Kay P, Berlin B, Maffi L, Merrifield W et al. (1997). Color naming across languages.
  In: Hardin C, Maffi L (eds), Color categories in thought and language.
  Cambridge University Press, pp 21–58
- Kay P, Maffi L (1999) Color appearance and the emergence and evolution of basic color lexicons. Am Anthropol 101(4):743–760
- Kuehni RG (2001) Determination of unique hues using Munsell color chips. Color Res Appl 26(1):61–66
- Labov W (1990) The intersection of sex and social class in the course of linguistic change. Lang Var Change 2(2):205–254
- Lindsey DT, Brown AM (2002) Color naming and the phototoxic effects of sunlight on the eye. Psychol Sci 13(6):506–512
- Lindsey DT, Brown AM (2004) Sunlight and "blue": the prevalence of poor lexical color discrimination within the "grue" range. Psychol Sci 15(4):291–294
- Lindsey DT, Brown AM (2006) Universality of color names Proc Natl Acad Sci USA 103(44):16608–16613
- Lindsey DT, Brown AM (2009) World Color Survey color naming reveals universal motifs and their within-language diversity Proc Natl Acad Sci USA 106 (47):19785–19790
- Lindsey DT, Brown AM (2014) The color lexicon of American English. J Vis 14 (2):17–17
- MacDonald L, Mylonas D (2014) Gender differences for colour naming in Spanish and English. In: Ortiz G, Ortiz C, R. R eds Color, culture and identity: past, present and future. AIC. AMEXINC, Oaxaca, pp 422–427
- Miller DG (2014) English lexicogenesis. OUP, Oxford
- Milroy J, Milroy L (1993) Mechanisms of change in urban dialects: the role of class, social network and gender. Int J Appl Linguist 3(1):57–77
- Moore C, Romney AK, Hsia T-L (2002) Cultural, gender, and individual differences in perceptual and semantic structures of basic colors in Chinese and English. J Cognit Cult 2(1):1–28
- Murray IJ, Parry NR, McKeefry DJ, Panorgias A (2012) Sex-related differences in peripheral human color vision: a color matching study. J Vis 12(1):18–18
- Mylonas D, Paramei GV, MacDonald L (2014) Gender differences in colour naming. In: Anderson W, Biggam CP, Hough CA, Kay CJ eds Colour studies: a broad spectrum. John Benjamins, Amsterdam/Philadelphia, pp 225–239
- Nevalainen T (2000) Gender differences in the evolution of Standard English: evidence from the Corpus of Early English Correspondence. J Engl Linguist 28(1):38–59
- Nowaczyk RH (1982) Sex-related differences in the color lexicon. Lang Speech 25 (3):257-265
- Paramei GV, Griber YA, Mylonas D (2018) An online color naming experiment in Russian using Munsell color samples. Color Res Appl 43(3):358–374
- Pardo PJ, Perez A, Suero M (2007) An example of sex-linked color vision differences. Color Res Appl 32(6):433–439
- Pérez-Carpinell J, Baldovi R, de Fez MD, Castro J (1998) Color memory matching: time effect and other factors. Color Res Appl 23(4):234–247
- Regier T, Kay P, Cook RS(2005) Focal colors are universal after all Proc Natl Acad Sci USA 102(23):8386–8391
- Rodríguez-Carmona M, Sharpe LT, Harlow JA, Barbur JL (2008) Sex-related differences in chromatic sensitivity. Vis Neurosci 25(3):433–440
- Samarina LV (2007) Gender, age, and descriptive color terminology in some Caucasus cultures. In Maclaury RE, Paramei GV, Dedrick D (eds), Anthropology of color: interdisciplinary mulfilevel modeling. John Benjamins Publishing Company, Amsterdam/Philadelphia, pp 457–466
- Saucier DM, Elias LJ, Nylen K (2002) Are colours special? An examination of the female advantage for speeded colour naming. Pers Individ Differ 32(1):27–35
- Saunders B (2000) Revisiting basic color terms. J R Anthropol Inst 6(1):81–99
  Shen X (2005) Sex differences in perceptual processing: performance on the colorKanji stroop task of visual stimuli. Int J Neurosci 115(12):1631–1641
- Simpson J, Tarrant AW (1991) Sex-and age-related differences in colour vocabulary Lang Speech 34(1):57–62

Swinkels I (2015) Unusual features in the colour classification of Modern Irish. Master's thesis, Leiden University

Szeszel M, Alvarado N, Jameson K, Sayim B (2005) Semantic and perceptual representations of color: evidence of a shared color-naming function. J Cognition Cult 5(3–4):427–486

Thomas LL, Curtis AT, Bolton R (1978) Sex differences in elicited color lexicon size. Percept Mot Skills 47(1):77-78

Uusküla M, Bimler D (2016) From listing data to semantic maps: cross-linguistic commonalities in cognitive representation of colour. Folk: Electron J Folk 64:57–90. http://www.folklore.ee/folklore/vol64/colour.pdf

Volbrecht VJ, Nerger JL, Harlow CE (1997) The bimodality of unique green revisited. Vis Res 37(4):407–416

Yang Y (1996) Sex- and level-related differences in the Chinese color lexicon. Word 47(2):207–220

Yang Y (2001) Sex and language proficiency level in color-naming performance: an ESL/EFL perspective. Int J Appl Linguist 11(2):238–256

#### **Acknowledgements**

Parts of this work were supported by the NSF grant SMA-1416907.

#### **Author contributions**

NLK and NAF designed the study, performed data analysis, and wrote the paper. All authors reviewed the manuscript.

#### **Competing interests**

The authors declare no competing interests.

### **Additional information**

**Supplementary information** is available for this paper at https://doi.org/10.1057/s41599-019-0341-7.

Correspondence and requests for materials should be addressed to N.A.F. or N.L.K.

Reprints and permission information is available at http://www.nature.com/reprints

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

CC ① Op

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing,

adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <a href="http://creativecommons.org/licenses/by/4.0/">http://creativecommons.org/licenses/by/4.0/</a>.

© The Author(s) 2019