DOI: 10.2352/ISSN.2470-1173.2016.16HVEI-114

Individual Differences in Color Naming

Delwin T. Lindsey [1,2], Angela M. Brown[2], [1]Department of Psychology, Ohio State University; Mansfield, OH. USA; [2] College of Optometry, Ohio State University, Columbus, OH, USA.

Abstract

Color naming shows great diversity worldwide, across languages and even within languages, but there are only about 11 named color categories, which are universal across languages and occur in only a few distinct, universal motifs. Most world languages contain multiple motifs among their speakers. Color communication using these motifs is not optimal, even when controlling for the number of color terms they contain. This is due to the diversity across individuals, including: which colors are not named, which motif is used, low consensus about the terms for the colors that are named, and where the boundaries are between the colors.

Introduction

The evolution of color terms has been the topic of research for at least 150 years, starting when scholars in Europe began to study languages spoken from people on other continents. Investigators and scholars have articulated three competing hypotheses about how the observed diversity in color naming has come about. These are the "universalist" hypothesis, developed by Paul Kay and his collaborators [1, 2], "linguistic relativity", also known as the "Sapir-Whorf" hypothesis [3], and the "emergence hypothesis," which is a more recent version of linguistic relativity [4].

According to Berlin & Kay [1], languages partition the set of all discriminable colors exhaustively into at least two but no more than about 11 color categories, each with its own Basic Color Term. These 11 categories are universal across the 96 languages they examined, but any given language names a subset of them. Berlin & Kay observed that the universal color categories occur in only a limited number of combinations within languages. The structure of these combinations suggested to them that color lexicons evolve over time along a constrained trajectory, and every language presently spoken is at some stage along this trajectory. At first, all colors were divided into only two or three named color categories. Over time, more color terms were added to the lexicon, in quasi-fixed order, by partitioning the existing color categories into smaller and smaller parcels, until, eventually, the 11 universal Basic Color Terms came to be used. According to Berlin & Kay, every modern language is at some particular position along a seven-stage trajectory of color term evolution. For example, most speakers of Yacouba, spoken in Africa, have only *black*, *white* and *red* in their lexicons, and are at Stage II, whereas speakers of English, who are at Stage VII, use all 11 Basic Color Terms: *black*, *white*, *red*, *yellow*, *green*, *blue*, *brown*, *orange*, *pink*, *purple*, and *gray*.

Other investigators have proposed other accounts of the differences between languages in how colors are named. Some investigators account for the differences across languages by the different roles that colors play in the cultural lives of the people that speak them. According to these "linguistic relativists" [5], colors are free to vary *ad libitum* across languages, depending on the needs of the people to name the colors of important items in their cultural or natural environment.

One particular version of the linguistic relativity explanation, known as the "emergence hypothesis [4]," holds that ancestral languages named colors using terms that referred to culturally significant items of that color. For example, in English, the color term *orange* names the color using the name of the fruit. The emergence hypothesis predicts that ancestral color terms named only the few colors that are similar to corresponding items of the same name in the environment, leaving many, culturally insignificant colors unnamed. Thus, the emergence hypothesis stands in clear contrast to the universalist hypothesis of Berlin & Kay, in that Berlin & Kay predict that every color should have a name at every stage of color term evolution, but Levinson predicts that early color-naming systems should be sparse, leaving large regions or color space without corresponding color terms.

The World Color Survey

The World Color Survey ("WCS", collected under the direction of Paul Kay and his colleagues [2]) is a large database of color naming data provided by 2616 informants. About 24 speakers of each of 110 unwritten languages, spoken in pre-industrial cultures, provided a single color term for each of 330 Munsell color samples (Fig. 1A, circled color samples). In their analysis of that data set, Kay et al. found evidence of all 11 Basic Color Terms, but they also found evidence of a more complex pattern of color term evolution than previously, with several sub-stages and versions of the seven stages originally proposed by Berlin & Kay.

A

B

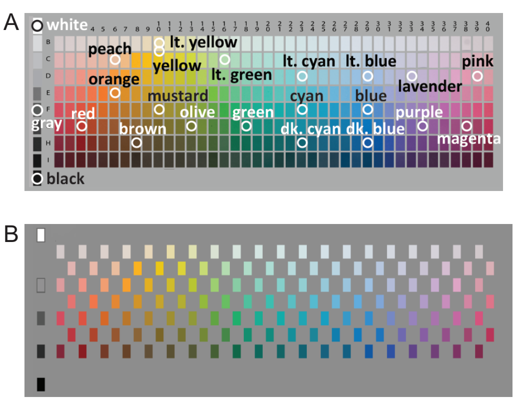

Figure 1. A, The stimuli of the WCS, shown in their order within the Munsell color order system, with the 40 hues in columns and eight values in rows and 10 neutrals on the side. The circled colors are the subset of the WCS stimuli that were used in our study of the Hadza of Tanzania. B, the subset of 145 samples used in our large study of the Somali people living in Columbus, OH.

We performed a cluster analysis of the WCS data set [6], which revealed 11 universal color terms that were remarkably similar to Berlin & Kay's Basic Color Terms. We refer to these as "glossed" color terms because the individual words in different languages (e.g., *red*, *guduud*, and *tisiuneya* in English, Somali, and

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.16HVEI-114

Hadzane, respectively, three non-WCS languages that are discussed below) and within the same language (e.g., *red*, and *scarlet* in English) are all counted as the same term. Then we performed a second cluster analysis of the individual informants' usage of those universal glossed color terms. The second cluster analysis provided evidence of only about four universal color naming systems (we call them "motifs"), within which different combinations of color terms are used [7]. Remarkably, these motifs recur, with little variation, across WCS languages, including languages with no known historical linguistic or cultural ties, and spoken on different, widely separated continents (Fig. 2). The motifs differed from each other mostly in how they named the cool colors, so we named them after their cool color terms. The motifs were: "Dark" (*black*, *white*, and *red*), "Gray" ("*black*, *white*, *yellow*, *gray*, and *red*), "Grue" (*black*, *white*, *red*, *yellow* and *grue*), and "Green-Blue" (*black*, *white*, *red*, *yellow*, *green*, and *blue*, as well as, sometimes, other colors). Here, "*grue*" is defined as a single color term that names both green and blue samples.

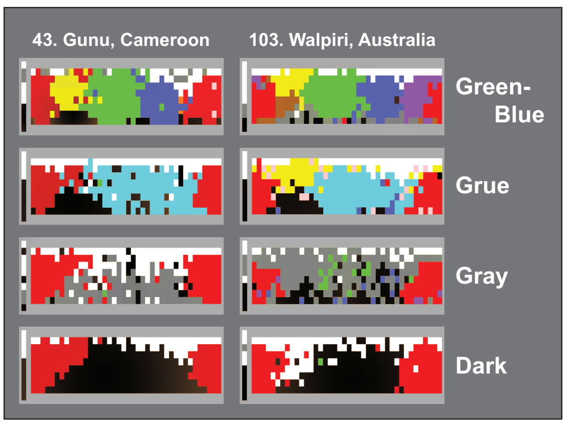

Figure 2. Color naming data from four individuals naming colors in Gunu and Walpiri (languages #43 and 104 in the WCS). Although these two cultures are on different continents, and their languages have no known historical linguistic relation the similarities across the two languages and the similar variation within the languages, are striking. The diagrams show color samples, positioned as in Fig. 1A, false-colored to indicate the color terms used by the informants. Cyan codes "grue", the color term that covers both blue and green color categories.

A salient feature of WCS color lexicons, which is not accounted easily for by the framework of Berlin & Kay [1] and Kay et al. [2] is the striking individual differences in color naming observed across the speakers of most languages of the WCS [7-10]. Much of this individual variability is due to the remarkable fact that multiple motifs are observed among informants speaking most languages: the modal number is three of the four motifs in a given language [7]. For example, in Fig. 2, all four motifs are represented among the speakers of Gunu, (spoken in Cameroon) and also among the speakers of Warlpiri (spoken in Australia).

Although some investigators view these differences in color naming as "empirical noise" [11], we believe that these differences provide important insight into how color lexicons form and evolve. We have argued that the distribution of the motifs across the speakers of a language suggests that color lexicons evolve by changing the number of speakers that use each of the motifs [7]. Thus, a language might evolve from being mostly a "Grue" language to being mostly a "Green-Blue" language by changing the fraction of speakers who use the Grue and Green-Blue motifs.

Mutual information and color communication.

How well do people actually communicate about color? It seems intuitively obvious that speakers of a language with only two or three color terms should not be able to communicate very well, no matter how those color terms map onto the set of colors. It seems equally obvious that a language containing a single motif with more color terms should allow its speakers to communicate better. To evaluate how well people communicate about color with others speaking the same language, we performed an information theoretic analysis of the color naming data from the WCS.

We define the quality of color communication between two informants as the value of Mutual Information [12] resulting from a color communication game. We illustrate the calculation of mutual information by considering one informant, the "sender" (S), who views a set of N color samples and tries to communicate the colors of each of those samples to a second informant, the "receiver" (R). In this example, N=23 samples. S randomly chooses one sample, with replacement, and names the chosen sample in his or her language. R views his/her own duplicate set of 23 samples, and tries to identify the sample that was chosen and named by S. Of course, if S and R do not speak the same language, R will have no way of understanding which sample S intended to communicate, so R will have a 1/23 chance of guessing correctly. The entropy associated with the uncertainty in color identification for $i = 1 \ldots 23$ samples will be:

$$-\sum_i (\tfrac{1}{23})_i log_2((\tfrac{1}{23})_i) = 4.52 \text{ bits}. \quad (1)$$

However, if S and R do speak the same language, R's chances are improved. Even if S says "don't know", R can choose a sample that R does not know the word for, and R's performance will likely be above chance. Mutual Information is a measure of the amount by which S's utterance improves R's chances of identifying S's chosen color sample. Letting $I(C_R; C_S)$ be the reduction in uncertainty (here, the reduction from 4.52 bits) in R's identification of the test samples C_R, given the utterances by S associated with the samples C_S,

$$I(C_R; C_S) = \sum_{s,r} p(s,r) log_2\left(\frac{p(s,r)}{p(s)p(r)}\right). \quad (2)$$

p(s,r) is an N x N matrix (a 23 x 23 matrix in our example) of the joint probability distribution on the random variables C_S and C_R (Fig. 3). The entries in the cells of this matrix are the probabilities associated with the s = red, black, white, green, … samples that S names, and the r = red, black, white, green, … samples selected by R in response to S's names. *p(r)* and *p(s)* are the marginal distributions on C_R and C_S, respectively, that is, they are the sums of the columns and rows, respectively. Thus *p(s,r)* explicitly represents the probability of R's color selections, given S's intended sample; that is, *p(s,r)* is based on color communication from S to R.

To evaluate the mutual information across a language community, we tabulate the outcomes for all $N_L = n(n\text{-}1)$ pairwise permutations of n informants speaking language L. Then we aggregate the results of the N individual color communication games (up to now, we've been using N=23). The result is the Group Mutual Information (GMI), which is our measure of the rate of color communication:

$$GMI = I_N(C_R; C_S) = \sum_{s,r} p_N(s,r) log_2\left(\frac{p_N(s,r)}{p_N(s)p_N(r)}\right). \quad (3)$$

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.16HVEI-114

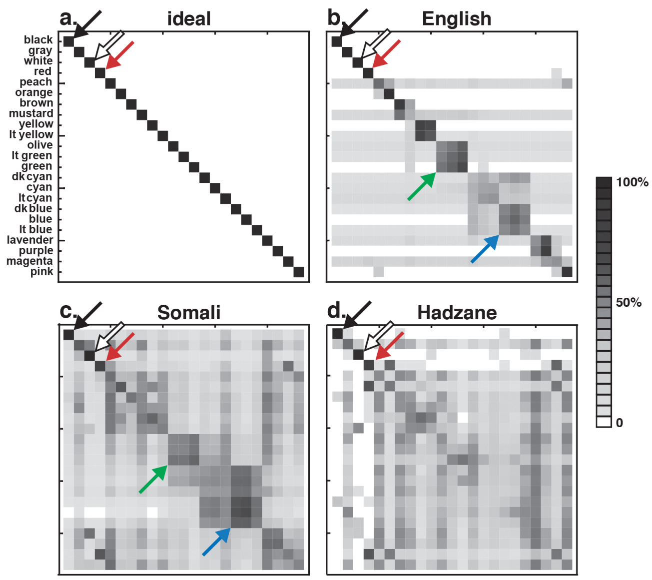

Figure 3. Joint probability distribution matrix diagrams. Graytones indicate p(s,r), the fraction of informants choosing a particular sample (the columns) based on the color terms uttered by the senders (rows: messages sent by S; conventional English names of the colors on the left of Panel A). Downward arrows: black, white and red. Upward arrows: aggregations of density at green and blue. A, the ideal situation where every informant uses a unique name for each sample, which is understood by every other informant. B, English data. The concentration of density near the minor diagonal indicates that most "receivers" understood what most "senders" were saying. The low density rows of coloring arise when the receiver had to guess what was meant. C., Somali data; D, Hadzane data. In C, D, dark colors concentrated along the minor diagonal indicate effective color communication: the receiver understood the message of the sender. Color off the minor diagonals indicate color communication failures: the receiver failed to choose the color sample communicated by the sender.

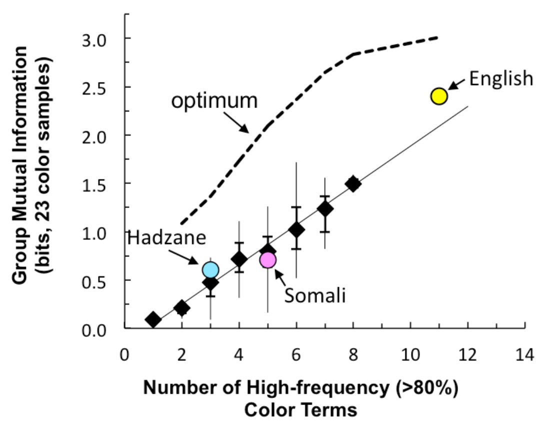

Figure 4. Group Mutual Information (Equation 3) as a function of the number of high-frequency Basic Color Terms in the WCS languages (diamonds: medians, interquartiles and full ranges), based on the 23-sample subset of the WCS stimulus set. Disks, color naming in the three languages discussed here, based on data collected with the 23-sample stimulus set [13]. See text for further details. Dashed line, optimum performance for each number of color terms.

Median values of GMI for the 110 languages in the WCS appear as diamonds in Fig. 4. It is not surprising that there is a positive association between the number of color terms a language and the mutual information communicated among its speakers. That association is highly statistically significant ($r=0.790$, $p<10^{-6}$), but it only accounts for 0.62 of the variance in the rate of color communication.

Color communication among Hadza, Somali, and U.S. informants

In collaboration with colleagues from the University of Pennsylvania, we have recently studied color naming in Hadzane [13], the language spoken by the Hadza, a Tanzanian hunter-gatherer society [14]. The stimulus set was 23 colors selected from the 330-sample stimulus set of the WCS (white circles in Fig. 1A). The stimulus set included 11 samples that WCS and English-speaking participants chose as the "best examples" of the Basic Color Terms of Berlin & Kay, plus 12 additional colors to sample the rest of the stimulus set. Unlike the WCS, we allowed informants to respond with "don't know" if they did not wish to name a particular color.

Hadza informants generally named *black*, *white* and *red* ("BWR") stimuli with high consensus, but individual informants named only a minority of the non-BWR colors, making generous use of the "don't know" option (there are many empty cells in Fig. 5A). Thus, Hadza color-naming was "sparse". Every informant named a different subset of the non-BWR samples, and no Hadza individual named all the WCS categories. However, most color categories in the stimulus set that are named in the WCS (or in English) were given a name by at least some Hadza (each column in Fig. 5a has at least one colored cell). Thus, Hadza color naming was "distributed" across the Hadzane data set. This shows that there is a surprisingly modern color naming system distributed across the collective responses of the Hadzane-speaking community, even though this modern system of color naming was not apparent in the responses of any individual Hadza informant. One could say that our task of finding the Hadzane color terms was "crowd-sourced" across the Hadza community.

Because of the high consensus for terms for the BWR colors and the low frequency and consensus in naming the non-BWR colors, we classify Hadzane as a three-term language. The interesting question is whether the evidence for the more modern color naming capability revealed by the non-BWR color terms increases the rate of color communication above that observed for other three-term languages in the WCS. In fact, it does not: the rate of color communication among Hadza is almost exactly on the regression line for the WCS (Fig. 4).

We have also collected data on monolingual speakers of Somali who have immigrated to Columbus, OH as refugees, and on U.S. undergraduate university students. For one data set [13], we used a method identical to the methods we used with the Hadza: we used the same 23-sample subset of the WCS stimuli, and we allowed "don't know" as a response. Somali informants used "don't know," but they did so less frequently than the Hadza informants (Fig. 5b), and Somali consensus for the terms they did name was higher than it was for the Hadza informants. The U.S. informants rarely used "don't know," and their consensus levels were nearly perfect for many of the samples in the stimulus set. Based on the colors named with high consensus, we place the Somali data set at 5 color terms, and the U.S data set at 11 color terms. The estimated rate of communication within those groups of informants was also very close to the regression line established for the WCS (colored disks in Fig. 4).

We also collected data without allowing "don't know," and many more samples: a 145-sample subset of the WCS stimuli for the Somali informants [15] (Fig. 1B) and the full 330-sample stimulus set (Fig. 1A) for the U.S. undergraduates.

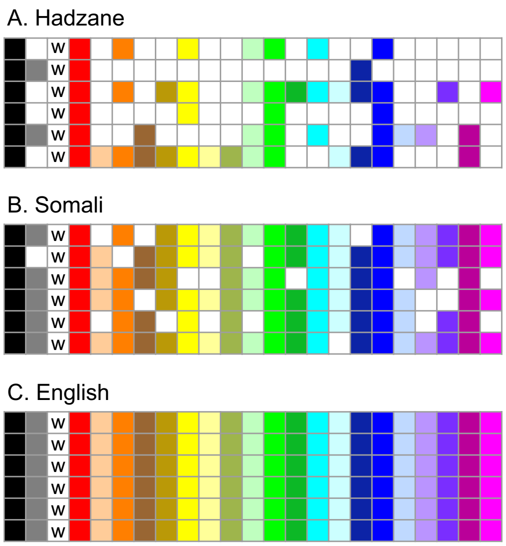

Figure 5. Representative examples of individual color naming data by Hadza and Somali informants. Each row is a different individual; the squares are the samples. False colors correspond to the sample colors named by the informants (white samples indicated with "w"), blank cells are "don't know" responses. All informants named the black, white and red samples. Each Hadza informant named a different subset of the remaining samples, with the samples corresponding to all the basic color terms represented among the responses of 51 informants. Somali informants named most samples, and every U.S. informant named all the samples.

Figure 3 shows grayscale plots of the group joint probability distribution matrices that were used to calculate the group mutual information values, across the three languages (Hadzane, Somali, and English) from the data collected using 23 samples. The rows of each matrix are the color samples named by the sender; the columns are, in order, the same set of colors that the receiver selected from. In Fig. 3, colors are arranged roughly in hue order: first, the achromatic colors (black, gray, white), then red, peach, orange, yellow…, … magenta, and pink.

Figure 3a shows the results of a communication game involving informants speaking a hypothetical language who use a different name for each of the 23 colors we used for the Hadza study. Consensus in the use of the 23 terms is 100%. In that hypothetical language, all non-zero values in the joint probability distribution matrix are maximal (1/23) and fall on the minor diagonal of the matrix. Similar matrices for English, Somali, and Hadzane show varying amounts of density off the minor diagonal, indicating failures of color communication. In English and especially Somali, there are large square areas of density in the areas of the green and especially the blue samples (green and blue upward-pointing arrows), indicating only approximate ability to communicate accurately the colors within those regions: a range of green and a range of blue samples receive the same names. Hadzane shows little density on the minor diagonal, except for *black*, *white* and *red* (arrows), indiating that only the BWR colors are named with high consensus and understood by all.

Optimal color communication

It is instructive to predict how well people using a given number of color terms could possibly communicate about the 23 color samples we studied. The prediction for optimal communication (from equation 3) is based on the assumption that consensus among informants was perfect. For example, if S and R agreed that there were 4 *cagaar* samples in the stimulus set, and if S called a particular sample *cagaar*, then R would choose an *cagaar* sample and would be correct with probability ¼.

The results of that analysis are the dashed curve labeled "optimum" in Fig. 4, based on the matrix shown in Fig. 3A. By this standard, the speakers of all three languages that we studied, as well as the speakers of all the WCS languages, fall far short of the optimal rate of communication for the number of color terms in their lexicons. This shortfall may be related to failure to meet the requirements for best communication, which are:

(1) Every speaker of a language should name every color in the stimulus set. In fact, "don't know" was a common response in our Hadza data set, and we suspect would have been very prevalent in the WCS, had it been and allowed response.

(2) Every speaker of the language should name the same number of color categories. In fact, the number of color categories varies greatly, as different individuals use different motifs.

(3) There should be perfect consensus about the color terms, across individuals. In fact, there is rarely perfect consensus even about the color categories where they agree on the range of colors associated with them.

(4) Speakers of a language should agree about where the boundaries between the color categories are. In fact, there is often poor consensus about the boundary locations across individuals, even within motifs.

(5) There should be equal numbers of samples assigned to each color category. In fact, color terms vary greatly in the number of samples they name.

Some of these points are speculative; others can be addressed from our data. For example, issue (1) is not very important, because the permissibility of "don't know" as a response had surprisingly little effect. Our English-speaking undergraduates rarely used "don't know", yet their data fall far short of the ideal performance for 11 color terms. Furthermore, when we compare our Somali informants' data when "don't know" was allowed [13] and when it was not allowed [16], on comparable sets of 23 samples, there was even a small decrement in performance in the "don't know" data sets (Fig. 6).

We examined the impact of issue (2) by analyzing our Somali data as a whole and separately by motif. The Somali informants in the main data set in ref. [16] used all four WCS motifs, with 17 informants using the Green-Blue motif and 6 using the Grue motif (we neglect the Gray and Dark motifs because there were only a few informants in each of those groups). Those who used the Green-Blue motif used more color terms and communicated better than those who used the Grue motif and fewer color terms. However, the rate of communication was still not much closer to the optimum than the data set as a whole. The speakers who used the Grue motif had one fewer color terms, but that also did not bring the optimum much closer to their performance. Thus, the existence of multiple motifs was not the main reason for the sub-optimal rate of color communication in Somali.

To address issue (3), we eliminated the differences across speakers in the actual color terms they used by glossing all the data sets to a single vocabulary. This manipulation did indeed improve

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.16HVEI-114

the GMI value, but again, not to the optimum value for the number of high-consensus terms in the lexicon.

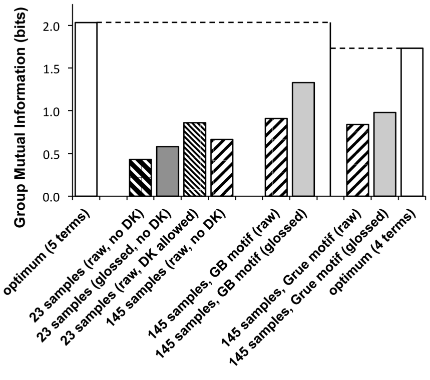

Figure 6. Group Mutual Information (Equation 3) for Somali participants, analyzed as discussed in the text. White bars, optimum performance. The Somali data set as a whole, and the GB motif, have 5 high-consensus color terms; the Grue motif has 4 high-consensus color terms. Hatched bars, raw data, with no glossing analysis. Gray bars, color terms glossed by cluster analysis. Glossing the terms into categories with single color terms, and restricting the analysis to single motifs both improved the GMI, but these manipulations, even in combination, do not bring the rate of communication up to its ideal value.

This leaves us with the irreducible facts that (4) participants often disagree as to where the (sometimes fuzzy) boundaries of the color categories are located in color space, and (5) the color categories in a language vary greatly in size, so the partition of colors space is never optimum for the number of color categories a person names. It is noteworthy that two groups of investigators have examined the optimality of the partition of color space [17, 18], and neither of them proposes that the color samples in the WCs stimulus set should be divided equally among categories. Indeed, same-sized color categories would have to be relative to the stimulus set used, so what is optimum for the analysis of data collected using the WCS stimulus set would not necessarily be optimum for whatever set of colors the person might see daily.

For these reasons, the rate of communication, calculated from the color naming data from our Hadza, Somali, and U.S. informants, as well as the color naming data from the WCS, fall far short of what should be theoretically possible, even when taking into account the number of high-consensus color terms in those languages. All of these issues except issue (5) are traceable to the individual variation in the naming of colors. Thus, individual differences in color naming are not merely noise: they are, in fact, at the heart of our understanding of language-mediated communication among individuals.

Discussion

The regularity of individual differences in color naming within and across cultures suggest that these differences are not noise. Rather, they reflect fundamental processes important to color categorization and color term acquisition and evolution. We argue that individual variability in how colors are named and understood can provide critical data for understanding the origins and evolution of lexical representations of color.

Most investigators, regardless of which hypothesis they subscribe to, agree that ancient color naming systems included few terms, and that more terms were added until about 11 color categories were named. Thus, color lexicons are thought to have evolved from simpler to more complex. Because the Hadza live purely by hunting and gathering, much as ancient humans lived, it is interesting to speculate that the color naming system in the Hadzane language might be similar to that of our ancestors. The Hadza color lexicon is both simple, because it contains only three frequently-used color terms, and also complex, because the variability across individual reveals the existence of many color terms that are found in languages like English, which is spoken in modern, highly industrialized society. Thus, color naming by the Hadza as a group does not easily fit into a single position along the simple-to-complex trajectory that is at the heart of most thinking about color naming system evolution.

Our results from the WCS and on Hadza, Somali, and U.S. observers are consistent with a universal representation of color. There are indeed only about 11 universal color categories, across all these languages, including languages that have no known historical or linguistic ties. Furthermore, the motifs into which these color categories are organized occur with little variation worldwide. However, the prevalence of multiple motifs among the lexicons of individuals within so many language communities, including those we have studied personally, is not easily reconciled with the view that each language is at a particular stage along a universal trajectory of color term evolution. Finally, the sparseness of our Hadzane data (Fig. 5) challenges the view that all colors are nameable in all languages.

The linguistic relativity hypothesis holds that color categories are largely free to vary across languages, and usage of color terms in a language is determined by the cultural and technological needs of the society where the language is spoken. Most proponents of the linguistic relativity hypothesis agree that the color terms are constrained by the contiguity of colors within categories, but this constraint leaves great leeway across languages as to how color categories should be established and named. The fact that the WCS data set revealed 11 universal color terms and four universal motifs, all of which recurred across unrelated languages worldwide, does not obviously support the view that there are few constraints how languages name colors. The fact that individuals within cultures use different motifs is not consistent with the hypothesis that the particular culture where a language is spoken is the only determinant of what its named color categories are: something besides the shared cultural experience must be responsible for the individual variability with each culture.

The emergence hypothesis, which is a version of the linguistic relativity hypothesis, has several features that agree remarkably well with our results. The central idea of the emergence hypothesis is that ancestral lexical representations of color were sparse because the color terms were named after items in the cultural or the natural environment, whose colors were located in specific regions of color space. As color lexicons evolved, these specific named colors generalized to wider regions of color space, until eventually all the colors were named. Unlike other versions of the linguistic relativity hypothesis, this view offers an explicit account of how culture determines what groups of colors will be named. The emergence hypothesis is nicely consistent with high frequency of “don’t know” responses provided by Hadza informants who tried to name non-BWR samples, because the emergence hypothesis predicts that colors that do not correspond to items in the environment should be un-named. However, many non-BWR color categories are named in Hadzane, although the frequency of naming and consensus for these color terms is low. It is difficult to evaluate the emergence hypothesis as a full account of color naming in Hadzane, because Hadzane is not currently well enough

DOI: 10.2352/ISSN.2470-1173.2016.16HVEI-114

studied linguistically to determine whether the Hadzane color terms are in fact names of things in the Hadza environment. Moreover, we find excellent agreement between the named color categories in Hadzane and the universal WCS color categories, which suggests that the color categories are not entirely determined by the unique features of the Hadza lifestyle. However, there is room here for at least some environmental universals, such as the colors of charcoal and blood, which might partly rescue the emergence hypothesis in the face of cross-cultural correspondence in color terms.

Conclusions

In this report, we reviewed two groups of studies of the naming of colors, which we compared to the predictions of three standard accounts of human color naming behavior. We have shown that world color terms and the motifs in which they occur are universal across the 110 world languages in the WCS [2], which is not obviously consistent with the great diversity across languages predicted by the linguistic relativity hypothesis [5]. Furthermore, there is great diversity within languages, in that most languages include multiple motifs across the idiolects of their speakers. This diversity presents an additional challenge to the linguistic relativity hypothesis, in that something besides the shared experience unique to a culture determines the color lexicon of the speakers of a language. We have also shown that at least one language, Hadzane, shows very sparse color naming among its speakers. Although most Hadza name more colors than just BWR (*black*, *white*, and *red*), no one names all the colors, and the consensus for the non-BWR colors is low. This sparseness of color naming indicates that the universalist notion that every language partitions all visible colors exhaustively into named color categories is not correct. Future research, in which "don't know" is a permitted response, will be required to determine whether the partition structure is common or rare among languages spoken in pre-industrialized societies. The emergence hypothesis suggests that few languages that have small numbers of color terms will show this partition structure.

Color terms are for color communication. Color communication effectiveness is necessarily well short of optimal, because optimal color communication would require a distinct color term for every color, and perfect consensus as to what that color term should be. Human language distinguishes only a modest number of color categories (certainly less than 20 categories, perhaps no more than 11 categories), even though human beings can distinguish up to about a million distinct colors visually. Our work on human color naming indicates that people do not even communicate as well as their 11 color categories could allow them to do. This failure of communication is the result of individual variation in color naming behavior.

Acknowledgements

This work was supported by NSF BCS-1152841. We express our thanks to Coren Apicella and David Brainard, of the University of Pennsylvania, who collaborated with us on the Hadza project.

References

1. Berlin, B., and Kay, P. (1969). Basic Color Terms: Their Universality and Evolution, (Berkeley and Los Angeles: University of California Press).
2. Kay, P., Berlin, B., Maffi, L., Merrifield, W.R., and Cook, R. (2010). The World Color Survey, (Berkeley: CSLI Publications).
3. Lucy, J.A. (1997). The linguistics of color. In Color Categories in Thought and Language, C.L. Hardin and L. Maffi, eds. (Cambridge: Cambridge University Press).
4. Levinson, S.C. (2000). Yélî Dnye and the theory of basic color terms. Journal of Linguistic Anthropology *10*.
5. Lucy, J.A. (1997). Linguistic Relativity. Annual Review of Anthropology *26*, 291-312.
6. Lindsey, D.T., and Brown, A.M. (2006). Universality of color names. Proceedings of the National Academy of Sciences of the United States of America *103*, 16608-16613.
7. Lindsey, D.T., and Brown, A.M. (2009). World color survey color naming reveals universal motifs and their within-language diversity. Proceedings of the National Academy of Sciences of the United States of America *206*, 19785 -- 19790.
8. Rosch, E. (1972). Probabilities, sampling, and ethnographic method: The case of Dani colour names. Man, New Series *7*, 448-466.
9. Webster, M.A., and Kay, P. (2007). Individual and population differences in focal colors. In Anthropology of Color: Interdisciplinary and multilevel modeling, R.E. MacLaury, G.V. Paramei and D. Dedrick, eds. (Amsterdam: John Benjamins), pp. 29-54.
10. Webster, M.A., and Kay, P. (2012). Color categories and color appearance. Cognition *122*, 375-392.
11. Regier, T., Kemp, C., and Kay, P. (2015). Word Meanings across Languages Support Efficient Communication. The handbook of language emergence *87*, 237.
12. Cover, T.M., and Thomas, J.A. (1991). Elements of Information Theory, (Hoboken, N. J.: Wiley).
13. Lindsey, D.T., Brown, A.M., Brainard, D.H., and Apicella, C.L. (2015). Hunter-Gatherer Color Naming Provides New Insight into the Evolution of Color Terms. Current Biology.
14. Marlowe, F. (2010). The Hadza: Hunter-gatherers of Tanzania, (Univ. of Calif. Press).
15. Brown, A.M., and Lindsey, D.T. (2015). The color lexicon of the Somali language. Journal of Vision *in press*.
16. Brown, A.M., Isse, A., and Lindsey, D.T. (2016). The color lexicon of the Somali langauge. Journal of Vision *in press*.
17. Jameson, K.A., and D'Andrade, R.G. (1987). It's not really red, green, yellow, blue: An inquiry into perceptual collor space. In Color Categories in Thought and Language, C.L. Hardin and L. Maffi, eds. (Cambridge: Cambridge University Press), pp. 295–319.
18. Regier, T., Kay, P., and 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*, 1436-1441.

Author Biographies

Delwin T. Lindsey received his BA in physics from Pomona College (1971) and his PhD in biopsychology from the University of Chicago (1985). Since then he has worked in the departments of psychology at the University of Washington and the Ohio State University, where he is now professor, and in the College of Optometry at the Ohio State University. His work has focused on human color vision.

Angela M. Brown received her MS in experimental psychology from Rutgers University (1974) and her PhD in experimental psychology from the University of Michigan (1981). Since then she has worked at Bell Telephone Laboratories, at the University of Washington, and in the colleges of optometry at SUNY and Ohio State University, where she is now professor. Her work has focused on human visual development and human color vision