Color naming across languages reflects color use

Edward Gibson[a,2], Richard Futrell[a], Julian Jara-Ettinger[a,1], Kyle Mahowald[a], Leon Bergen[a], Sivalogeswaran Ratnasingam[b], Mitchell Gibson[a], Steven T. Piantadosi[c], and Bevil R. Conway[b,2]

[a]Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139; [b]Laboratory of Sensorimotor Research, National Eye Institute and National Institute of Mental Health, National Institutes of Health, Bethesda, MD 20892; and [c]Department of Brain and Cognitive Sciences, University of Rochester, Rochester, NY 14627

Edited by Dale Purves, Duke University, Durham, NC, and approved August 15, 2017 (received for review November 30, 2016)

What determines how languages categorize colors? We analyzed results of the World Color Survey (WCS) of 110 languages to show that despite gross differences across languages, communication of chromatic chips is always better for warm colors (yellows/reds) than cool colors (blues/greens). We present an analysis of color statistics in a large databank of natural images curated by human observers for salient objects and show that objects tend to have warm rather than cool colors. These results suggest that the cross-linguistic similarity in color-naming efficiency reflects colors of universal usefulness and provide an account of a principle (color use) that governs how color categories come about. We show that potential methodological issues with the WCS do not corrupt information-theoretic analyses, by collecting original data using two extreme versions of the color-naming task, in three groups: the Tsimane', a remote Amazonian hunter-gatherer isolate; Bolivian-Spanish speakers; and English speakers. These data also enabled us to test another prediction of the color-usefulness hypothesis: that differences in color categorization between languages are caused by differences in overall usefulness of color to a culture. In support, we found that color naming among Tsimane' had relatively low communicative efficiency, and the Tsimane' were less likely to use color terms when describing familiar objects. Color-naming among Tsimane' was boosted when naming artificially colored objects compared with natural objects, suggesting that industrialization promotes color usefulness.

color categorization | information theory | color cognition | Whorfian hypothesis | basic color terms

The question of color-naming systems has long been caught in the cross-fire between universality and cultural relativism. Cross-cultural studies of color naming appear to indicate that color categories are universal (1–3). However, the variability in color category boundaries among languages (4), and the lack of consensus of the forces that drive purported universal color categories (5, 6), promotes the idea that color categories are not universal, but shaped by culture (7). Here, we focus on two color categories, WARM and COOL, which are not part of the canonical set of "basic" categories proposed by Berlin and Kay but which nonetheless may be fundamental (8, 9), and which might relate to the basic white/black categories occupying the first stage in the Berlin/Kay hierarchy. Are WARM and COOL universal categories, and if so, why? Here we address these questions by collecting an extensive original dataset in three cultures and leveraging an information-theoretic analysis that has been useful in uncovering the structure of communication systems (10, 11).

The World Color Survey (WCS) provides extensive data on color naming by 110 language groups (www1.icsi.berkeley.edu/wcs/data.html). However, the WCS instructions are complex, and different WCS researchers likely adopted different methods (*SI Appendix*, section 1 and, Figs. S1 and S2), possibly undermining the validity of the WCS data (12, 13). To evaluate this possibility, we obtained extensive color-labeling data using two extreme versions of how the WCS instructions might have been implemented: a "free-choice" paradigm that placed no restrictions on how participants could name colors, and a "fixed-choice" paradigm on separate participants, where participants were constrained to only say the most common terms we obtained in the free-choice paradigm. We conducted experiments in three groups: the Tsimane' people, an indigenous nonindustrialized Amazonian group consisting of about 6,000 people from lowland Bolivia who live by farming, hunting, and foraging for subsistence (14); English speakers in the United States; and Bolivian-Spanish speakers in Bolivia, neighboring the Tsimane'. Tsimane' is one of two languages in an isolate language family (Mosetenan). Although there is trade with local Bolivian towns, most of the Tsimane' participants have limited knowledge of Spanish. To the extent that the communities have organized schooling, education is conducted in Tsimane'.

We analyzed our data using information theory, building on work that suggests that color naming can be better understood by considering informativeness rather than opponent-process theory (3, 11, 15). This analysis can be understood in terms of a communication game (16–18). Imagine that a speaker has a particular color chip c in mind and uses a word w to indicate it. The listener has to correctly guess c, given w. On each trial, the listener guesses that c is among a set of the chips; the listener can pick a set of any size and is told "yes" or "no." The average number of guesses an optimal listener would take to home in on the exact color chip provides a measure of the listener's average surprisal (S, measured in bits; Eq. **1**), a quantitative metric of communication efficiency. The surprisal score for each color c is computed by summing together a score for each word w that might have been used to label c, which is calculated by multiplying $P(w|c)$ by $-\log(P(c|w)$, the listener's surprisal that w

Significance

The number of color terms varies drastically across languages. Yet despite these differences, certain terms (e.g., red) are prevalent, which has been attributed to perceptual salience. This work provides evidence for an alternative hypothesis: The use of color terms depends on communicative needs. Across languages, from the hunter-gatherer Tsimane' people of the Amazon to students in Boston, warm colors are communicated more efficiently than cool colors. This cross-linguistic pattern reflects the color statistics of the world: Objects (what we talk about) are typically warm-colored, and backgrounds are cool-colored. Communicative needs also explain why the number of color terms varies across languages: Cultures vary in how useful color is. Industrialization, which creates objects distinguishable solely based on color, increases color usefulness.

Author contributions: E.G., S.T.P., and B.R.C. designed research; E.G., R.F., J.J.-E., S.R., M.G., and B.R.C. performed research; E.G., R.F., J.J.-E., K.M., L.B., S.R., and B.R.C. analyzed data; and E.G. and B.R.C. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]Present address: Department of Psychology, Yale University, New Haven, CT 06520.

[2]To whom correspondence may be addressed. Email: bevil@nih.gov or egibson@mit.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1619666114/-/DCSupplemental.

would label c. We estimate $P(c|w)$ via Bayes Theorem assuming a uniform prior on $P(c)$.

$$S(c) = \sum_{w} P(w|c)\log\frac{1}{P(c|w)}. \qquad [1]$$

The novelty of our analysis is to define communicative efficiency about a particular color chip, relative to the set of color chips, so that we have an information-theoretic measure—average surprisal—that allows us to rank the colors for their relative communicative efficiency within a language. To compare color communication efficiency across languages, we follow others (3, 11) in estimating the overall informativeness of the color system of each language by averaging the average surprisal across the chips (Eq. **2**; see *SI Appendix*, section 3 for a worked-out example):

$$\sum_{c} P(c)S(c). \qquad [2]$$

Results

Color Naming. In our first analysis, we determined the most frequently used (modal) term for each color chip. Identifying modal terms within a language provides an objective way to relate one language to another, but modal terms are not a good measure of the sophistication of a language because they do not account for the variability in term use within a population. Compared with Bolivian-Spanish and English speakers, the Tsimane' speakers showed much greater variability in what color terms were used for all chromatic chips except red. Participants also showed high consensus in labeling black and white chips; we did not solicit responses for other achromatic Munsell chips, although some of the 80 colored chips were labeled using terms glossed as "black" or "white." Diamond sizes in Fig. 1 show the fraction of participants who reported the modal term for the chip at each location in the color array in the free-choice task; diamonds are smaller, on average, in Tsimane' compared with English and Bolivian-Spanish, indicating higher variability in color-term use among the Tsimane' (individual-participant results are shown in *SI Appendix*, Figs. S3–S5; *SI Appendix*, Table S3 shows the key which links the color of the diamonds to the modal color terms). Across the 80 chips, the Tsimane' had eight modal color terms whereas English had 10 and Bolivian-Spanish had 11 (*SI Appendix*, Table S3). Despite differences in variability of term use among the languages, modal term assignments resulted in a generally similar partitioning of the color space across all three languages we studied (similar results were obtained using the fixed-choice version of the task; *SI Appendix*, Fig. S6). These results show that, at the population level, all three languages have a comparable representation of color space. The results for the Tsimane' people mirror observations made in the Hadzane of Tanzania, another indigenous community previously thought to have limited color knowledge, but now known to possess a rich color lexicon distributed across the population (3).

Objects from Memory. We were concerned that the higher variability among the Tsimane' might reflect unfamiliarity with the Munsell cards. To test this possibility, we asked the participants to tell us what color word they associated with familiar objects (a test of memory color). Given each object (O), we determined the uncertainty (H, simple entropy) over color words (W) that were used to refer to that object (Eq. **3**).

$$H(W|O) = -\sum_{i} P(W_i|O)\log(P(W_i|O)). \qquad [3]$$

Eq. **3** quantifies across people in the culture the consistency of the color label for a given object. Although the objects were selected because each has a characteristic color and is familiar to Tsimane' speakers, Tsimane' on average had higher uncertainty over the

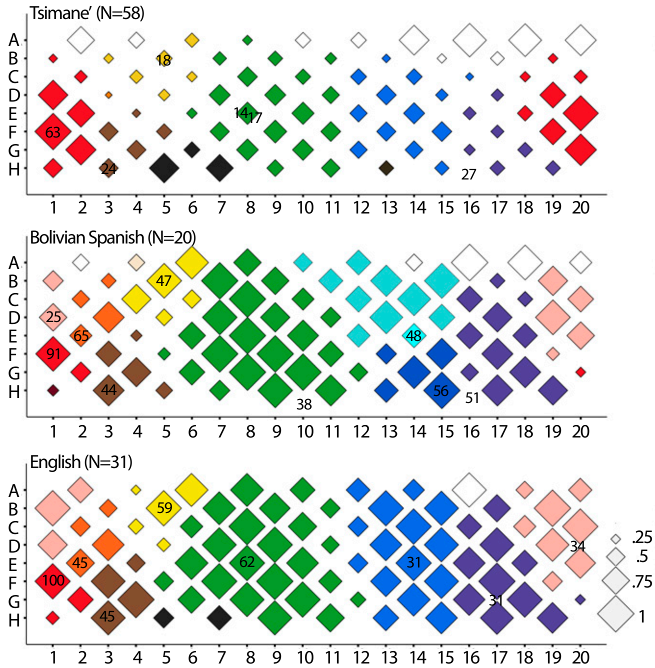

Fig. 1. The Amazonian Tsimane' people show large individual differences in color naming, but at the population level, similar color categories to those observed among Bolivian-Spanish and English speakers. Color naming of 80 chips evenly sampling the Munsell array, presented singly in random sequence under controlled lighting, in Tsimane', Spanish-speakers in neighboring San Borja (Bolivia), and English-speaking students near Boston (see *SI Appendix*, Table S2 for the key relating the axes to Munsell chip designations). Color of each diamond corresponds to the modal color for the chip (see *SI Appendix*, Table S3 for the key matching the color with the terms in each language). Diamond size shows the fraction of people who gave the modal response. All participants showed 100% consistency for black and white chips: negro, blanco (Bolivian-Spanish); tsincus, jaibas (Tsimane'). The location of the numbers overlying the plot indicate the color chips in the 160-chip Munsell array that were most frequently selected as the best example of the subset of modal color terms queried (*SI Appendix*, Table S4). The numbers are the percentage of respondents who made the given selection. Note that two modal color terms in Tsimane', *yụshñus* and *shandyes*, correspond to the same chip (E8). English speakers were asked about red, green, yellow, blue, orange, brown, purple, and pink. Bolivian-Spanish speakers were asked about rojo, verde, amarillo, azul, celeste, anaranjado, morado, cafe, and rosa. Tsimane' were asked about jäinäs (~red), yụshñus (~blue), shandyes (~green), itsidyeisi (~purple), cafedyeisi (~brown), and chames (~yellow). Data are from the free-choice version of the task (n = 58 Tsimane', 20 Bolivian-Spanish, 31 English); data from the fixed-choice version of the task, conducted in separate participants (n = 41 Tsimane', 25 Bolivian-Spanish, 29 English), yielded similar results (*SI Appendix*, Fig. S6).

color words they associated with the objects (1.06 bits); uncertainty was comparable between English (0.33 bits) and Bolivian-Spanish (0.30 bits) [in paired t tests between the entropy scores per object, the Tsimane'-English comparison is $t(15) = -2.88$, $P = 0.01$; the Tsimane'-Spanish comparison is $t(15) = -3.16$, $P = 0.006$; the English-Spanish comparison is $t(16) = -0.16$, $P = 0.9$]. This difference was driven predominantly by objects whose color names have generally low agreement among Tsimane' as assessed in the color-chip naming experiment (yellow, orange, green, blue, brown), but high agreement among English and Bolivian-Spanish speakers (Fig. 2). Objects such as blood, hair, and teeth whose colors fall in high-consensus color categories in Tsimane' (glossed red, black, and white) were associated with low uncertainty (all Tsimane' speakers have black hair). The relatively large variability in memory colors associated with familiar objects among the Tsimane' corroborates the conclusions drawn from the naming of color chips. Additional control experiments assessing reaction times

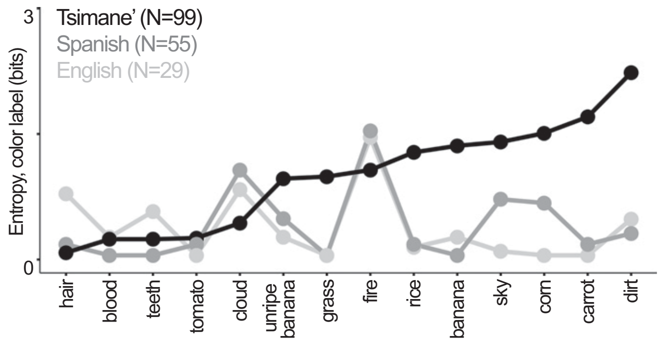

Fig. 2. Variability of color labels (entropy, Eq. **3**) for familiar objects, ordered by Tsimane' results. On average, Tsimane' has higher entropy over color words for a particular object (1.06 bits, compared with English, 0.33 bits, and Bolivian-Spanish, 0.30 bits).

further supports the conclusion that participants were fully engaged in the various tasks (*SI Appendix*, section 2 and Fig. S7).

Similar Results for Free-Choice vs. Fixed-Choice Versions of Color-Naming Task. Returning to the analyses of the color-chip naming tasks, the higher color-naming variability among the Tsimane' speakers, compared with English and Bolivian-Spanish speakers, is reflected in higher average surprisal for all color chips, using either open (*SI Appendix*, section 3 and Fig. S8) or fixed versions of the task. Averaging across color chips (Eq. **2**), the Tsimane' color system (4.88 bits) was shown to be less informative than that of English (3.80 bits) or Bolivian-Spanish (3.86 bits) (free-choice task). Unsurprisingly, the average number of color words produced in each population during the fixed-choice task was lower than in the free-choice version (Fig. 3*A*). Remarkably, the overall informativeness of a language was very similar for the two versions of the task (Fig. 3*A* and *SI Appendix*, section 4 and Figs. S9–S11; the tasks were performed in separate sessions, with different people, about a year apart). Furthermore, Spearman correlations of the rank-ordered sequence of color chips (ordered by increasing average surprisal) for each version of the task were high (Tsimane': $\rho = 0.71$; Bolivian-Spanish: $\rho = 0.90$; English: $\rho = 0.92$). These results come as a relief, allaying widespread concerns about the methodology of the WCS, and licensing further information-theoretic analyses of the WCS data (12, 13).

Focal Color Analyses and Unique Hues. In separate experiments, for several frequently used color terms in each language, subjects were asked to indicate which color chip was the best representative of the color term. Such "focal" colors can be reasonably predicted by statistical models that identify the most representative color chip given each speaker's color-naming data (19). The contours in Fig. S8 show the probability density of color samples chosen for red, green, blue, yellow (English); rojo, verde, azul, amarillo (Bolivian-Spanish); jainas, shandyes, yushñus, chamus (Tsimane') (*SI Appendix*, section 5 and Table S4 provides the chip designations for the most frequently chosen chip for each focal color). The contours tend to cover a broader area of the array for Tsimane' speakers compared with English or Bolivian-Spanish speakers, consistent with the results of the other color-naming experiments showing that the Tsimane' are more variable in the color terms they use (Fig. 1). The contours in *SI Appendix*, Fig. S8 are for colors that correspond to the "unique hues," which have long been postulated to be psychological primary colors (20): The unique hues are considered to be "irreducible" primaries which cannot be described using any more primitive labels (unlike "orange," which some consider yellowish red). Given their purported primary status, one might have hypothesized that these colors would be associated with relatively low average surprisal. Contrary to this prediction, we found that only the red and yellow focal colors had low surprisal across all three languages. The relatively low surprisal of red and yellow, compared with the higher surprisal of blue and green, recalls the smaller individual differences in unique hue settings for red and yellow compared with blue and green (21). These results add to a growing body of research suggesting that the unique hues might not be as special as widely thought (6, 22, 23). Instead, the results suggest that warm colors (reds, yellows) are associated with higher communicative efficiency compared with cool colors (blues, greens).

Analysis of Average Surprisal Values Within Each Language. The pattern of average surprisal values (the variations in gray shown in *SI Appendix*, Fig. S8) was consistent across the three languages. Consistent with this impression, the rank-ordered sequence of color chips was similar across the three languages (Fig. 3*B*; Spearman rank correlation, between Bolivian-Spanish and English $\rho = 0.87$; between Bolivian-Spanish and Tsimane' $\rho = 0.51$; between English and Tsimane' $\rho = 0.53$; Table S5; *SI Appendix*, section 6 and Fig. S12). The ordering forms a striking pattern that is not determined by the unique hues or the focal colors. Warm-colored chips (red, pink, orange, yellow, brown) across Tsimane', English, and Bolivian-Spanish showed relatively low average surprisal, whereas cooler colors (blues, greens) showed higher average surprisal. The rank ordering is also not explained by Berlin and Kay's proposed order of acquisition (1), which has blue and green arising before pink, orange, and brown. Our results suggest that despite overall gross differences in the communication efficiency across languages, among chromatic chips, warm colors are always the easiest to communicate precisely. Remarkably, we found that this relationship was true across the entire WCS of 110 languages (Fig. 4 and *SI Appendix*, section 7 and Fig. S13). These results provide an explanation for the universal distinction between warm and cool colors: Warm colors are always associated with higher communicative efficiency compared with cool colors. Together the results suggest two complementary conclusions: All languages, even those with very few consensus color terms, have a comprehensive

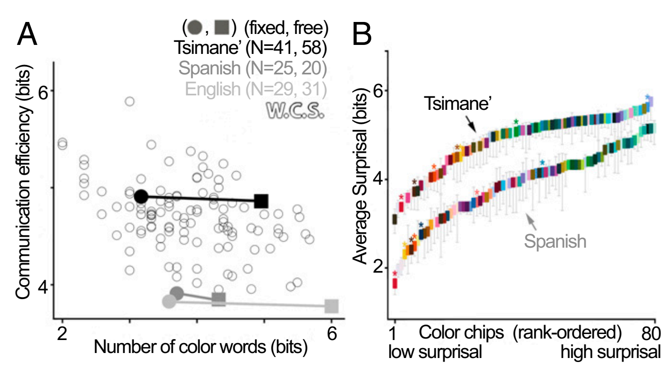

Fig. 3. Communication efficiency of color naming, across languages and among color chips. (*A*) Communication efficiency for each language of the WCS (open symbols), Tsimane' (black symbols), Bolivian-Spanish (dark gray symbols), and English (light gray symbols), as a function of number of unique color words used by the population of participants tested in each language. The two data sets collected in Tsimane', Bolivian-Spanish, and Tsimane' show that variability in experimental methods have little impact on assessments of communicative efficiency of color naming, licensing the use of the WCS data for further analysis. Circles show data from experiments in which participants were constrained to use a fixed vocabulary of basic color terms; squares show data where participants were free to use any term. Number of participants stated as (N=fixed choice, free choice). Communicative efficiency for each language was computed using Eq. **2**. (*B*) Color chips rank-ordered by their average surprisal (computed using Eq. **1**) for Tsimane' and Bolivian-Spanish (pattern for English overlaps Spanish, omitted for clarity). *SI Appendix*, Table S5 provides the chip identity in rank order. The asterisks represent focal colors determined as described in Fig. 2. The sequences of colors in each population are highly correlated (Spearman rank correlation between Bolivian-Spanish and English, $\rho = 0.87$; between Bolivian-Spanish and Tsimane', $\rho = 0.51$; and between English and Tsimane', $\rho = 0.53$).

PSYCHOLOGICAL AND COGNITIVE SCIENCES

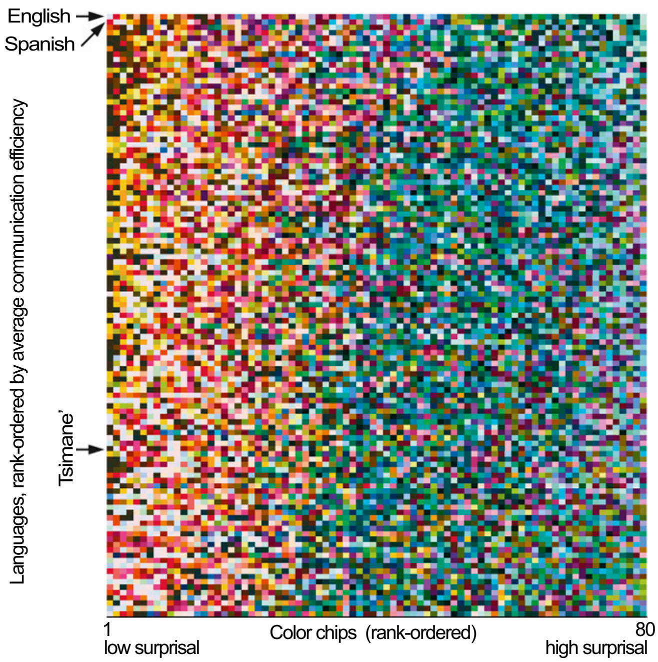

Fig. 4. Color chips rank-ordered by their average surprisal (computed using Eq. **1**) for all languages in the WCS, and the three languages tested here. Each row shows data for a given language, and the languages are ordered according to their overall communication efficiency (Eq. **2**).

color lexicon distributed across the population; however, all languages, even those with a very sophisticated color language, prioritize the same set of (warm) colors.

The colors of objects versus backgrounds. The discovery that warm colors are more precisely communicated compared with cool colors is a finding that emerges from the information-theoretic analysis. However, what determines this universal asymmetry in communicative efficiency between warm and cool colors? We hypothesized that the ordering of chips by average surprisal arises because of the color statistics associated with salient objects, in contrast to their indistinct backgrounds. Natural scenes typically do not show an equal representation of colors. Instead, warm colors (yellow/orange/red) and cool colors (blue/green) are overrepresented (24–26), regardless of season or ecosystem (27), and primary visual cortex is adapted to these statistics (28, 29). Attempts have been made to relate the chromatic statistics of a small sample of natural images to color categories (30), but it is not clear whether objects are representative of natural images in general. To fill this gap in knowledge, we analyzed the colors of objects identified by independent observers in a dataset of 20,000 photographs; this dataset was curated by Microsoft from over 200,000 photographs for the purpose of depicting salient objects (31). We discovered that pixel colors for the objects were more often within the red/yellow/orange ("warm") range, compared with backgrounds, which were typically blue/green ("cool"). Moreover, the likelihood that a color would be found in an object was negatively correlated with its average surprisal in the three languages we studied (Fig. 5 and *SI Appendix*, section 8 and Fig. S14) and the 110 languages of the WCS. These results suggest that what determines the universal patterns across the diversity of languages is the consistent link between warm colors and behaviorally relevant items—salient objects—in the environment. We confirmed these conclusions in an analysis of spectral measurements obtained from objects with and without behavioral relevance to trichromatic primates (32). We found that behaviorally relevant objects (such as fruit eaten by the animals) tended to have colors associated with lower average surprisal (*SI Appendix* and Fig. S15).

These results support the hypothesis that usefulness is the reason why languages acquire a color name. The relatively low communicative efficiency of color naming among Tsimane' suggests that color is simply not that useful for this population. The Tsimane' have little exposure to artificial (industrialized) objects. Industrialization has greatly expanded the gamut and color stability of objects. One idea is that exposure to artificially colored objects promotes the usefulness of color for object identification, which is hypothesized to promote greater precision in color language (33, 34). To test this idea, we performed a contrastive-labeling task (35) (*SI Appendix*, section 9). Eight pairs of objects, familiar to Tsimane' and English, were used in the experiment, four pairs of natural objects and four pairs of artificial objects. Participants were first presented with one object and were asked to name it. Then they were presented with the second object of the same type and asked to name it (Fig. 6*A*). Tsimane' were much less likely to use a color term (Fig. 6*B*). But a mixed-effect logistic regression showed a main effect of object class among the Tsimane': They were more likely to use a color word when naming artificial compared with natural objects ($\beta = 3.59$, $z = 4.00$, $P < 0.0001$), which is consistent with the idea that industrialization promotes color-naming efficiency.

Discussion

The debate on the origins of color categories pits the hypothesis that color-naming systems emerge from universal underlying principles determined innately against the view that culture determines color categories; it is often implied that only one or the other of these theories is correct. Our results favor a reconciliation of these ideas through the the efficient-communication hypothesis (11), which states that categories reflect a tradeoff between informativeness of the terms and their number (10). Cultures across the globe show common patterns in color naming, and even languages with few high consensus color terms appear to have a complete color lexicon distributed across the population, as shown by Lindsey et al. (3) for the Hadza of

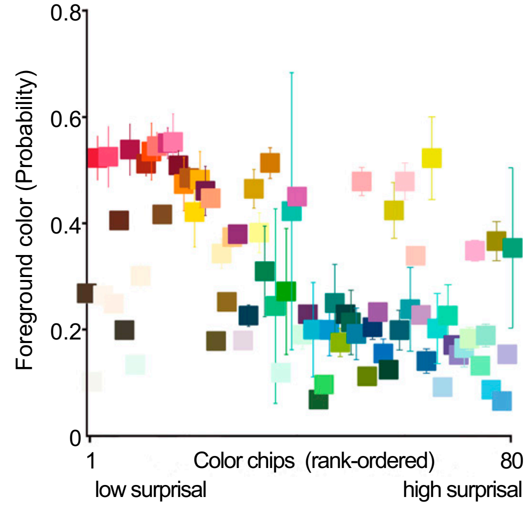

Fig. 5. The color statistics of objects predict the average surprisal of colors. Objects in the Microsoft Research Asia database of 20,000 photographs were identified by human observers who were blind to the purpose of our study (31). The colors of the pixels in the images were binned into the 80 colors defined by the Munsell chips used in the behavioral experiments (across the images there were 9.2×10^8 object pixels and 1.54×10^9 background pixels). The y axis shows the [(number of pixels of given color in objects)/(number of pixels of given color in objects + number of pixels of given color in backgrounds)]; the color chip ranking is that obtained for the Tsimane'. Error bars are SE. The three languages were not significantly different from each other (English: slope = −0.0064, $\rho = -0.57$, P value = 3×10^{-8}; Bolivian-Spanish: slope = −0.0049, $\rho = -0.44$, P value = 5×10^{-4}; Tsimane': slope = −0.0054, $\rho = -0.49$, P value = 5×10^{-6}).

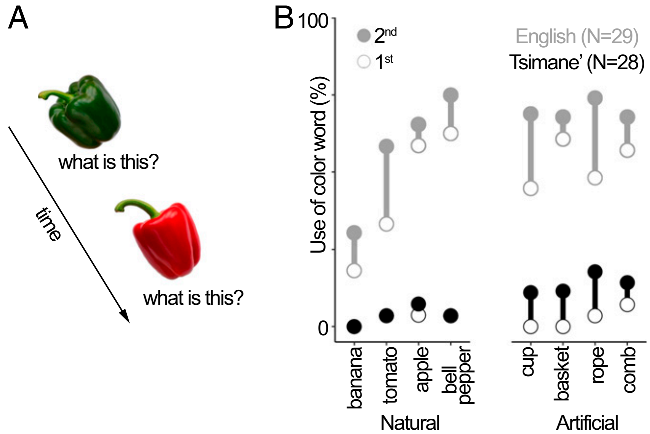

Fig. 6. The Tsimane' use color terms less frequently than English speakers. (*A*) Contrastive-labeling paradigm, adapted to assess use of color terms in normal communication. (*B*) Percent of trials in which participants used a color word to describe objects presented in sequential pairs. Members of each pair were identical except for color (e.g., green banana/yellow banana). Tsimane' speakers were less likely to use a color word (mixed effect logistic regression, $\beta = -5.23$, $z = -5.48$, $P < 0.0001$). Among Tsimane', a mixed-effect logistic regression shows a main effect of artificiality ($\beta = 3.59$, $z = 4.00$, $P < 0.0001$) and presentation order ($\beta = 1.57$, $z = 3.09$, $P < 0.01$) with no interaction ($\beta = 0.91$, $z = 1.19$, $P = 0.23$). Among English, we find a main effect of presentation order ($\beta = 1.53$, $z = 4.00$, $P < 0.001$).

Africa and by us in the Tsimane' of South America. These common patterns across cultures suggest some universal constraint on color naming, but the variability in communicative efficiency about color terms across cultures suggests that culture plays a role too. According to the communication-efficiency hypothesis, if a culture has little need for many high-consensus color categories, it is simpler in that communication system not to have them. We show that all cultures around the world favor communication about warm colors over cool colors, and that this phenomenon reflects a universal feature of natural scenes: Objects defined by human observers tend to be warm colored while backgrounds tend to be cool colored. These results provide evidence that usefulness is the reason for the addition of color terms (36, 37). For example, there simply are not that many natural blue objects, which may explain why many languages acquire the term "blue" relatively late (this left Homer scrambling to come up with an alternative description for the sea: "wine-dark" instead of "blue"; ref. 34). That many if not all "basic" color terms derive, historically, from the names of objects we care about (or cared about) provides yet another clue that usefulness is the principal force that drives color categorization. Consider "orange." Our results suggest that the color statistics of natural objects establish the relative salience of different colors and the informativeness of the associated terms. But we recognize that it is possible that the causal relationship is the inverse: that important natural objects acquired warm coloring to exploit the salience of these colors to trichromatic primates, for example to attract primates to assist in seed dispersal (32).

Although all languages appear to possess a fundamental distinction between warm and cool colors, the large variance of average surprisal values across languages suggests that the average usefulness of color varies among language groups. Our results on object/color associations support this hypothesis by showing that many objects that are common in both Tsimane' and US cultures have a diagnostic color term in English but not in Tsimane'. These results support the idea that color is not as useful for Tsimane' as it is for English and Bolivian-Spanish, consistent with findings in other non-Western groups (36). The Tsimane' have an extensive botanical vocabulary (14), which might obviate the need for color terms in their culture, which is heavily dependent on natural objects. Our results in a contrastive-naming task (Fig. 6) provide direct support for the idea that the predominance of artificially colored objects in Western cultures promotes the usefulness of color and, consequently, increases color-naming efficiency. The number of color terms used by Tsimane' individuals and the efficiency of color-term use increased with more exposure to Bolivian-Spanish (*SI Appendix*, section 10, Fig. S16, and Table S6), suggesting a mechanism for cultural transmission.

The present results confirm, for the Tsimane', prior work showing that language groups with relatively few consensus color categories nonetheless possess a large repertoire of color categories distributed across the population (3). The forces that give rise to the partitioning of color space into color categories more refined than warm/cool remain unclear, but our work promotes the idea the main driving force is the extent to which color categories are behaviorally relevant. Contributions to behavioral relevance may depend on stimulus saturation (38) and reflect efficient partitioning of the irregularities in perceptual color space (2, 15). Relative lightness must also be important in establishing behavioral relevance. Black was communicated among the Tsimane' with high efficiency, which replicates prior work showing that black and white are named reliably across all languages. The efficient communication of black is consistent with our overall hypothesis, that color categories reflect usefulness: Blacks are prevalent in natural images, and retinal processing favors darks over lights (39).

Color processing depends upon an extensive network of brain regions that process retinal signals (40), culminating in the highest levels of processing, in frontal cortex (41). The present report leverages color language as perhaps the best readout of this machinery as it pertains to behavior to uncover the forces behind the most fundamental color categorization, warm versus cool. Finally, we wonder to what extent the fundamental asymmetry in usefulness associated with warm colors versus cool colors underlies their emotional valence (42), as indicated by the warm/cool terminology itself.

Methods

All experimental procedures were approved by Massachusetts Institute of Technology's Committee on the Use of Humans as Experimental Subjects. Informed consent was obtained from all participants, as required by the Committee.

Color-Naming Munsell Chips. Participants were presented with each of 80 colored chips evenly sampling the standard Munsell array of colors (*SI Appendix*, Table S2) in a different random order for each participant under controlled lighting conditions using a light box (nine phosphor broadband D50 color-viewing system, model PDV-e, GTI Graphic Technology, Inc.). Each participant initially took a test of normal color vision (43). All participants that failed this task (~5% of participants) were excluded from further study. The task was performed indoors. For Tsimane' speakers, we assessed their knowledge of Spanish using a short questionnaire of very common words. Most participants did not know all of these words, suggesting a poor knowledge of Spanish for most (see *SI Appendix*, section 1 for more information).

Free-Choice Version. The instructions for this task were as follows: We want to know the words for colors in English/Spanish/Tsimane'. So we want you to tell us the colors of these cards. Tell us what other English/Spanish/Tsimane' speakers would typically call these cards. (Fixed-choice version of the task: There are 11 choices: black, white, red, green, blue, purple, brown, yellow, orange, pink, gray. Choose the closest color word.) See *SI Appendix*, section 1 for the Tsimane' and Spanish translations, with color terms from Spanish/Tsimane' for the fixed-choice version in *SI Appendix*, Table S3. Fifty-eight Tsimane'-speaking adults (mean age: 33.2 y; SD: 12.8 y; range 16–78; 38 females); 20 Spanish-speaking adults (mean age: 29.0 y; 9.1 SD: years; range 18–55; 11 females); and 31 English-speaking adults (mean age: 37.1 y; 11.6 SD: years; range 21–58; 10 females) completed this task. From the complete list of terms used in the population, we determined for each chip the term that was used most often (the modal term). Across the chips, we tallied the set of unique modal terms, and removed from the list any modal terms that were only used for one chip, thus omitting maracayeisi in Tsimane' (a color chip on which jäinäs, glossed red, was a close second), and fuschia and piel ("skin color") in Bolivian Spanish. This set of terms provides an estimate of the basic color terms in the population (*SI Appendix*, Table S3).

Fixed-Choice Version. Forty-one Tsimane' adults (mean age: 38.9 y; SD: 17.6 y; range 18–74; 24 females); 25 Spanish adults (mean age: 25.7 y; 9.1 SD: years;

range 18–55; 13 females); and 29 English adults (mean age: 26.0 y; 8.9 SD: years; range 18–55; 14 females) took part, where participants were given a fixed set of color labels to choose from for each color chip: the modal terms discovered in the free-choice version. We also included black/negro, white/blanco/gray/gris in the English/Spanish tasks, because they are regarded as basic color terms.

Focal Colors. Following the Munsell-chip color naming experiment, each participant ($n = 99$ Tsimane'; 55 Spanish; 29 English) was then presented with a standard 160-chip Munsell array of colors (illuminated by the lightbox) and was asked to point out the best example of several color words ("focal" colors). English speakers were asked about red, green, yellow, blue, orange, brown, purple, and pink. Spanish speakers were asked about rojo, verde, amarillo, azul, celeste, anaranjado, morado, cafe, and rosa. For Tsimane', in the free-choice version of the task, we asked about the colors that the participant produced. For the fixed-choice version, we asked about jäinäs (~red), yushñus (~blue), shandyes (~green), itsidyeisi (~purple), cafedyeisi (~brown), and chames (~yellow). The chips most often selected as focal colors for all of the terms probed are given in *SI Appendix*, Table S4. To show the population results and evaluate the possible privilege of the unique hues, we computed the probability density function for each of the four unique hues over the grid space. The contours in Fig. 2 show the probability that a given color word was used for each color chip on the basis of our empirical data.

Color-Naming Objects from Memory. Following the preceding two tasks, each participant ($n = 99$ Tsimane'; 55 Spanish; 29 English) was read a list of items that have typical colors, and was asked what color each item was in their experience. Each of the items had a conventional color in Tsimane' culture, usually the same as that in North American culture: a cloud (white/gray), dirt (brown), grass (green), hair (black), teeth (white), rice (white), an unripe banana (green), a ripe banana (yellow), the sky (blue), corn (yellow), yucca for eating (white), the outer husk of yucca (brown), blood (red), fire (orange/yellow), a carrot (orange), and a ripe tomato (red).

Spontaneous Use of Color in a Contrastive-Labeling Task. A subset of the participants who participated in the other experiments also took part in a contrastive-labeling experiment; some people took part only in the contrastive-labeling experiment and not in the Munsell-chip color-naming experiment. $n = 28$ Tsimane' adults (mean age: 30.9 y; SD: 17.8 y; range 18–90; 23 females) and 29 English participants (mean age: 35.5 y; SD: 11.0 y; range 21–58; nine females). Eight pairs of objects were obtained for naming, including four pairs of fruits and vegetables: a ripe (yellow) banana and an unripe (green) banana; a ripe (red) tomato and an unripe (green) tomato; a red apple and a green apple; a red bell pepper and a green bell pepper; and four pairs of artifact objects: a red and a yellow cup; a red and a blue comb; a red and a yellow piece of rope; and a red and a green small basket. All of the pairs of objects were identical except for their colors. Our method was an adaptation of the method used by Sedivy (35). Participants were first presented with one object and were asked to name the object. Then the second object of the same type was presented for naming. Each participant named all eight pairs of objects consecutively in this fashion. There were four different random orders of presentation. The experimenter/translator transcribed what was said.

ACKNOWLEDGMENTS. We thank Ricardo Godoy and Tomas Huanca for logistical help; Dino Nate Añez and Salomon Hiza Nate for help translating and running the task; Rashida Khudiyeva and Isabel Rayas for helping with the timing of the reaction times for the color and object naming; Serena Eastman and Theodros Haile who helped segmenting images for the natural-scene analysis; and Evelina Fedorenko, Alexander Rehding, Roger Levy, Caroline Jones, Joshua Green, and Neir Eshel for comments on the manuscript. The work was supported by NIH Grant R01EY023322 (to B.R.C.), NSF IOS Program Award 1353571 (to B.R.C.) and Linguistics Program Award 1534318 (to E.G.), and the Intramural Research Program of the NIH (National Eye Institute/National Institute of Mental Health). Additional support was provided by University of Rochester start-up funding (to S.T.P.) and the Wellesley College Undergraduate Research Summer Program.

1. Berlin B, Kay P (1969) *Basic Color Terms: Their Universality and Evolution* (Univ of California Press, Berkeley, CA).
2. Regier T, Kay P, Khetarpal N (2007) Color naming reflects optimal partitions of color space. *Proc Natl Acad Sci USA* 104:1436–1441.
3. Lindsey DT, Brown AM, Brainard DH, Apicella CL (2015) Hunter-gatherer color naming provides new insight into the evolution of color terms. *Curr Biol* 25:2441–2446.
4. Roberson D, Davidoff J, Davies IR, Shapiro LR (2005) Color categories: Evidence for the cultural relativity hypothesis. *Cognit Psychol* 50:378–411.
5. Valberg A (2001) Unique hues: An old problem for a new generation. *Vision Res* 41: 1645–1657.
6. Bohon KS, Hermann KL, Hansen T, Conway BR (2016) Representation of perceptual color space in macaque posterior inferior temporal cortex (the V4 complex). *eNeuro* 3:1–28.
7. Roberson D, Hanley JR (2007) Color vision: Color categories vary with language after all. *Curr Biol* 17:R605–R607.
8. Lindsey DT, Brown AM (2006) Universality of color names. *Proc Natl Acad Sci USA* 103: 16608–16613.
9. Holmes KJ, Regier T (2017) Categorical perception beyond the basic level: The case of warm and cool colors. *Cogn Sci* 41:1135–1147.
10. Piantadosi ST, Tily H, Gibson E (2011) Word lengths are optimized for efficient communication. *Proc Natl Acad Sci USA* 108:3526–3529.
11. Regier T, Kemp C, Kay P (2015) Word meanings across languages support efficient communication. *The Handbook of Language Emergence*, eds MacWhinney B, O'Grady W (Wiley-Blackwell, Hoboken, NJ), pp 237–263.
12. Lucy JA (1997) The linguistics of "color". *Color Categories in Thought and Language*, eds Hardin CL, Maffi L (Cambridge Univ Press, Cambridge, UK), pp 320–346.
13. Saunders BA, van Brakel J (1997) Are there nontrivial constraints on colour categorization? *Behav Brain Sci* 20:167–179, discussion 179–228.
14. Leonard WR, et al. (2015) The Tsimane' Amazonian Panel Study (TAPS): Nine years (2002-2010) of annual data available to the public. *Econ Hum Biol* 19:51–61.
15. Jameson K, D'Andrade RG (1997) It's not really red, green, yellow, blue: An inquiry into perceptual color space. *Hardin and Maffi Color Categories in Thought and Language* (Univ of Cambridge Press, Cambridge, UK), pp 295–319.
16. Lantz D, Stefflre V (1964) Language and cognition revisited. *J Abnorm Psychol* 69: 472–481.
17. Steels L, Belpaeme T (2005) Coordinating perceptually grounded categories through language: A case study for colour. *Behav Brain Sci* 28:469–489, discussion 489–529.
18. Baddeley R, Attewell D (2009) The relationship between language and the environment: Information theory shows why we have only three lightness terms. *Psychol Sci* 20:1100–1107.
19. Abbott JT, Griffiths TL, Regier T (2016) Focal colors across languages are representative members of color categories. *Proc Natl Acad Sci USA* 113:11178–11183.
20. Hering E *Outlines of a Theory of the Light Sense*, trans Hurvich LM, Jameson D (1964) (Harvard Univ Press, Cambridge, MA).
21. Webster MA, Miyahara E, Malkoc G, Raker VE (2000) Variations in normal color vision. II. Unique hues. *J Opt Soc Am A Opt Image Sci Vis* 17:1545–1555.
22. Bosten JM, Lawrance-Owen AJ (2014) No difference in variability of unique hue selections and binary hue selections. *J Opt Soc Am A Opt Image Sci Vis* 31:A357–A364.
23. Wool LE, et al. (2015) Salience of unique hues and implications for color theory. *J Vis* 15:10.
24. Vrhel MJ, Gerson R, Iwan LS (1994) Measurement and analysis of object reflectance spectra. *Color Res Appl* 19:4–9.
25. Nascimento SM, Ferreira FP, Foster DH (2002) Statistics of spatial cone-excitation ratios in natural scenes. *J Opt Soc Am A Opt Image Sci Vis* 19:1484–1490.
26. Webster MA, Mollon JD (1997) Adaptation and the color statistics of natural images. *Vision Res* 37:3283–3298.
27. Webster MA, Mizokami Y, Webster SM (2007) Seasonal variations in the color statistics of natural images. *Network* 18:213–233.
28. Conway BR (2001) Spatial structure of cone inputs to color cells in alert macaque primary visual cortex (V-1). *J Neurosci* 21:2768–2783.
29. Conway BR (2014) Color signals through dorsal and ventral visual pathways. *Vis Neurosci* 31:197–209.
30. Yendrikhovskij SN (2001) Computing color categories from statistics of natural images. *J Imaging Sci Technol* 45:409–441.
31. Liu T, et al. (2011) Learning to detect a salient object. *IEEE Trans Pattern Anal Mach Intell* 33:353–367.
32. Regan BC, et al. (2001) Fruits, foliage and the evolution of primate colour vision. *Philos Trans R Soc Lond B Biol Sci* 356:229–283.
33. Kay P, Maffi L (1999) Color appearance and the emergence and evolution of basic color lexicons. *Am Anthropol* 101:743–760.
34. Deutscher G (2010) *Through the Language Glass: Why the World Looks Different in Other Languages* (Metropolitan Books, New York).
35. Sedivy JC (2003) Pragmatic versus form-based accounts of referential contrast: Evidence for effects of informativity expectations. *J Psycholinguist Res* 32:3–23.
36. Kuschel R, Monberg T (1974) 'We don't talk much about colour here': A study of colour semantics on Bellona Island. *Man (Lond)* 9:213–242.
37. Levinson SC (2000) Yélî Dnye and the theory of basic color terms. *J Linguist Anthropol* 10:3–55.
38. Witzel C (2016) New insights into the evolution of color terms or an effect of saturation? *Iperception* 7:2041669516662040.
39. Ratliff CP, Borghuis BG, Kao YH, Sterling P, Balasubramanian V (2010) Retina is structured to process an excess of darkness in natural scenes. *Proc Natl Acad Sci USA* 107:17368–17373.
40. Lafer-Sousa R, Conway BR (2013) Parallel, multi-stage processing of colors, faces and shapes in macaque inferior temporal cortex. *Nat Neurosci* 16:1870–1878.
41. Bird CM, Berens SC, Horner AJ, Franklin A (2014) Categorical encoding of color in the brain. *Proc Natl Acad Sci USA* 111:4590–4595.
42. Palmer SE, Schloss KB, Sammartino J (2013) Visual aesthetics and human preference. *Annu Rev Psychol* 64:77–107.
43. Neitz M, Neitz J (2001) A new mass screening test for color-vision deficiencies in children. *Color Res Appl* 26:S239–S249.

Supplementary Information for Gibson et al. (2017) Color naming across languages reflects color use.

Supp. Materials, Methods, Analysis and Figures (SI-Section 1 to SI-Section 10; Figures S1-S16; Tables S1-S6)

Data collection with the Tsimane' was performed through daily trips to eight Tsimane' communities near San Borja, Bolivia, in collaboration with the Centro Boliviano de Investigación y de Desarrollo Socio Integral (CBIDSI).

SI-Section 1: Additional details of the Color-naming Task

The variability of the tasks that were run under the World Color Survey. All previous experiments in which participants from unindustrialized cultures were asked to label colors have used variants of the World Color Survey (WCS) instructions (1-3). These instructions introduce a complex notion of a "basic" color term, which takes several pages to describe. In writing these instructions, the authors of the WCS were trying to prohibit participants from producing low-frequency color terms like "scarlet" as a sub-class of red, or terms that are associated with only one object. The notion of "basic" color category does not include categories that are subsets of others, and can be applied broadly to many objects. But the concept of a basic color term has theoretical problems, because it is not clear that color categories cannot be parts of others, or that color categories cannot be very narrow; moreover, many languages simply do not have a super-ordinate concept of "color". Thus identifying "basic" color terms across languages begs the question of what counts as a basic color category (4).(4). The definition is also problematic in practice because it is so complex, making the notions difficult to explain, with the likely consequence that different WCS researchers implemented the complex instructions differently. An empirical evaluation of the WCS data suggests that there was variability in the kind of strategy that was used by WCS experimenters in implementing this task. The range of strategies can be captured by two extreme versions of the task: one in which participants could say whatever color words that they wanted – a "free-choice" version – and a second variant in which participants were restricted to choose a color word from a fixed set of choices – a "fixed-choice" version. For example, the fixed-choice version of the task was explicit when gathering the Pirahã WCS data, as discussed by Everett (5). Among the Pirahã queried in the WCS, all 25 participants except one produced all and only the same set of four words (one participant also used one additional word, in 5 trials); this outcome is extremely unlikely if the participants were not constrained to use a particular set of terms. We can compare Pirahã to the six other WCS languages which also have four modal color words. Two of these languages are like Pirahã, such that only the same four or five terms were provided by all of the participants. But participants in the other WCS languages with four modal color words produced more color terms: 15-17 terms in each of these four languages (sampling 25 people in each language). This corroborates the idea that WCS researchers may have used two versions of the task: a fixed-choice version (where only 4 words are used by all participants in these languages) and a free-choice version, with no such constraint, and the result that participants are much more variable in what they produce.

We quantified the variability in how the WCS task was implemented using two analyses. First, we examined the ratio of the total number of words that any participant used in a WCS language to the number of modal color terms. If a particular WCS task was implemented with a set of fixed choices for that language, this ratio will be close to one. But if there were fewer constraints on what words participants could use, then this ratio will result in a number larger than one. The histogram of the WCS ratios in **Figure S1** shows that many languages have a term-to-modal-term ratio of exactly one, suggesting a fixed-choice task in those languages. Some languages have a ratio very close to one, suggesting that some constraints were placed on what might be said in those languages. And many languages have much higher ratios, suggesting that no constraints were applied in these languages.

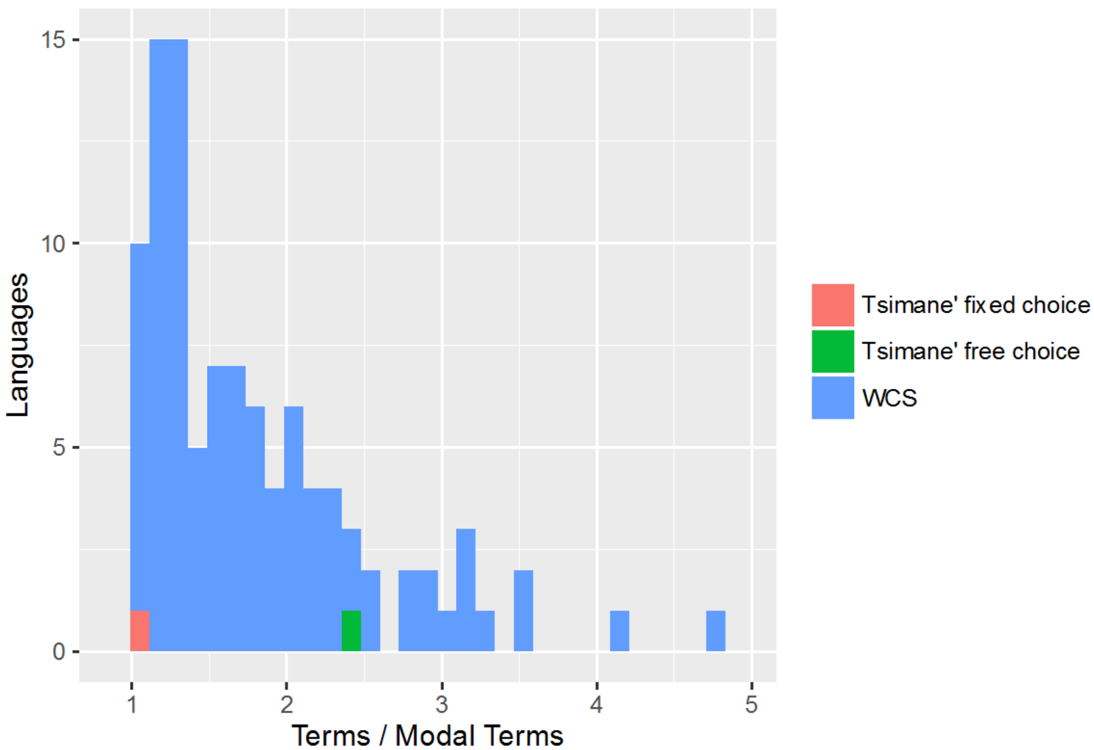

Figure S1. A histogram of the ratio of the number of words that any participant used in a WCS language to the number of modal color terms in that language. In this analysis, we restricted our attention to the subset of 80 color chips that were used in our experiments, in order to compare our results to those from the WCS. A ratio close to one suggests that the WCS task was implemented with a set of fixed choices for that language. Ratios that are much larger than one suggest that the WCS task was implemented with free choice of color terms for that language. We include the Tsimane' fixed-choice and free-choice ratios as baselines. For the bootstrap comparisons in the text, we compare only to the 99 WCS languages that have at least 20 participants. We randomly selected data from 20 Tsimane' subjects, and only include terms that appeared more than once (Tsimane' free choice = 18 total terms / 8 modal terms = 2.25).

What is the probability that we would observe each of the ratios in **Figure S1** if the task given to participants was to label colors freely? To answer this question, we calculated a distribution over term-to-modal-term ratios based on bootstrap resampling our Tsimane' free-choice data (see **Table S1**) for the 99 WCS languages that have at least 20 participants. This distribution tells us the probability that we would observe a certain term-to-modal-term ratio given randomly sampled subjects and a free-choice task. Most of the languages in the WCS dataset (80/99) have a term-to-modal-term ratio significantly less than the Tsimane' free-choice task, suggesting that these data were not collected with a fully free choice task. The data from the other 19 languages (those marked with "FALSE" in column 3 in **Table S1**) were plausibly generated with a fully free-choice task. Finally, seven of the 99 languages had term-to-modal ratios of exactly 1, suggesting that they were plausibly generated using the fixed-choice task.

Language	**term-to-modal-term ratio**	**Smaller than Tsimane' free-choice ratio? (p<.01)**
Abidji	1.33	TRUE
Agarabi	3.50	FALSE
Aguacateco	1.56	TRUE
Ampeeli	2.71	FALSE
Amuzgo	1.64	TRUE
Angaatiha	1.29	TRUE
Apinaye	1.83	TRUE
Arabela	1.86	TRUE
Bahinemo	1.29	TRUE
Bauzi	1.40	TRUE
Berik	2.67	FALSE
Bete	2.25	TRUE
Bhili	1.71	TRUE
Buglere	1.17	TRUE
Cakchiquel	1.64	TRUE
Camsa	1.73	TRUE
Carib	1.33	TRUE
Casiguran Agta	2.18	TRUE
CavineXa	1.17	TRUE
Cayapa	2.00	TRUE
Chcobo	1.00	TRUE
Chavacano	1.50	TRUE
Chayahuita	1.17	TRUE
Chinanteco	1.13	TRUE
Chiquitano	2.27	TRUE
Chumburu	1.88	TRUE
CofXn	1.00	TRUE
Colorado	1.20	TRUE
Culina	3.25	FALSE
Didinga	1.00	TRUE
Djuka	2.50	FALSE
Dyimini	1.43	TRUE
Eastern Cree	2.67	FALSE

Ejagam	1.00	TRUE
Ese Ejja	1.29	TRUE
Guahibo	1.30	TRUE
Guambiano	1.29	TRUE
Guarijio	1.83	TRUE
Gunu	3.00	FALSE
Halbi	2.75	FALSE
Huasteco	1.38	TRUE
Huave	1.20	TRUE
Iduna	3.40	FALSE
Ifugao	2.00	TRUE
Kalam	4.00	FALSE
Kamano-Kafe	2.86	FALSE
Kemtuik	2.14	TRUE
Kokoni	1.57	TRUE
Konkomba	2.80	FALSE
Kriol	1.30	TRUE
Kuku-Yalanji	2.40	TRUE
Kwerba	3.25	FALSE
Long-haired Kuna	2.11	TRUE
Mampruli	3.14	FALSE
Maring	2.43	TRUE
Martu Wangka	4.33	FALSE
Mawchi	1.29	TRUE
Mayoruna	1.00	TRUE
Mazahua	1.93	TRUE
Mazateco	1.30	TRUE
Menye	1.88	TRUE
Micmac	1.86	TRUE
Mikasuki	1.38	TRUE
Mixteco	1.50	TRUE
Murinbata	1.83	TRUE
Murle	1.57	TRUE
MXra PirahX	1.00	TRUE
Nafaanra	1.33	TRUE
NgXbere	2.29	TRUE
Ocaina	1.50	TRUE
Papago	2.00	TRUE
Patep	1.43	TRUE
Paya	1.40	TRUE
Saramaccan	2.18	TRUE

Sepik Iwam	1.80	TRUE
Seri	1.14	TRUE
Shipibo	1.25	TRUE
SirionX	2.00	TRUE
Slave	2.00	TRUE
Sursurunga	2.00	TRUE
Tabla	1.14	TRUE
Tboli	1.29	TRUE
Teribe	1.75	TRUE
Ticuna	1.33	TRUE
Tifal	3.20	FALSE
Tlapaneco	1.33	TRUE
Tucano	1.17	TRUE
Ucayali Campa	3.00	FALSE
Vagla	1.00	TRUE
Vasavi	1.50	TRUE
Walpiri	4.71	FALSE
Waorani	2.00	TRUE
Wobe	1.33	TRUE
Yacouba	1.00	TRUE
Yakan	1.09	TRUE
Yaminahua	1.80	TRUE
Yucuna	1.50	TRUE
Yupik	2.17	TRUE
Zapoteco	1.14	TRUE

Table S1. The term-to-modal-term for each of the 99 WCS languages with at least 20 participants, along with whether each ratio is significantly smaller than the ratio generated from samples of 20 participants in the Tsimane' free-choice task, at $p < 0.01$. When the ratio is significantly smaller, it provides evidence suggesting that the data from that language were not gathered using a fully free-choice task. The data from the other 19 languages (those marked with "FALSE" in column 3) were plausibly generated with a fully free-choice task.

Second, we examined the mean *color-word-overlap proportion* (CWO proportion) for the WCS languages, where the CWO proportion is defined as the mean proportion of color terms that each participant used which were also used by more than three-quarters of the other participants. A larger average CWO proportion for a language indicates a greater likelihood that words were constrained in the task. For example, 6 of the WCS languages have mean CWO proportions of 1.0, meaning that *every* term that a participant used was used by at least 75% of the other participants. Forty of the WCS languages have a CWO proportion of .9 or higher, suggesting a constrained vocabulary of color terms across participants, with few outlier terms. In contrast, there are 17 languages in the WCS with mean CWO proportions of 0.7 or below, meaning that 30% or more of the color terms that participants used in these languages were used by fewer than 75% of other participants. In these languages, there were probably no constraints on what speakers were told to say by their experimenters. Taken together, these two analyses suggest that the specific methods used to implement the WCS task were likely variable from one language to another.

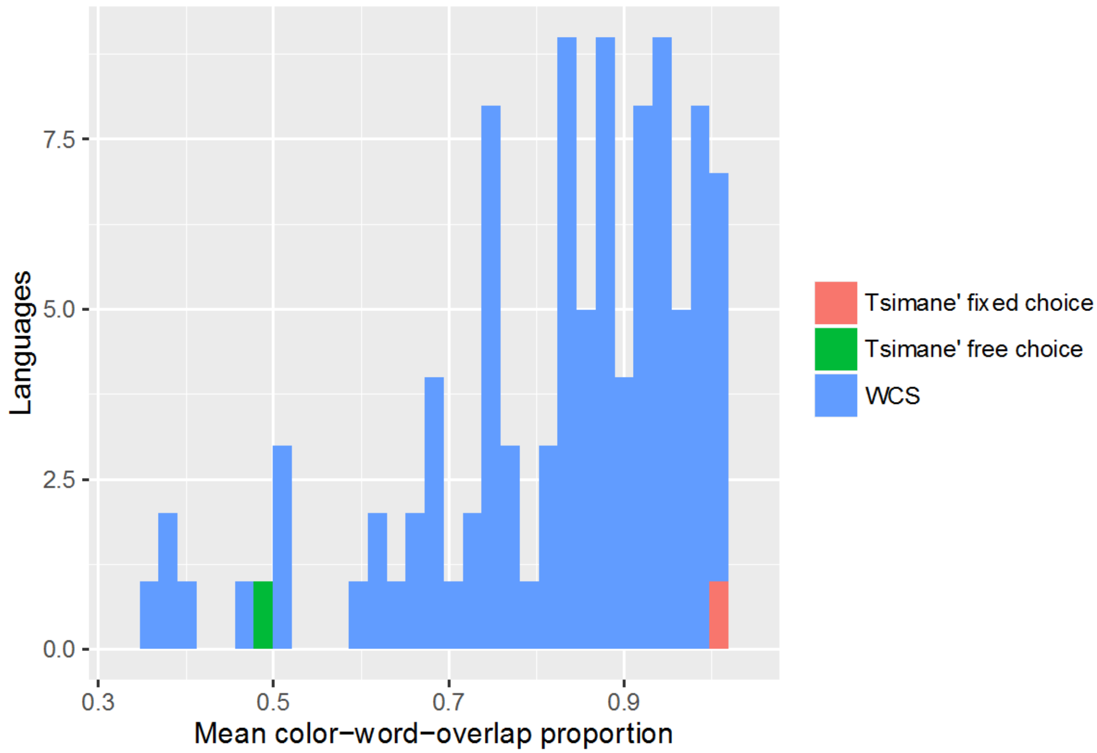

Figure S2. A histogram of the mean *color-word-overlap proportion* (CWO proportion) for the WCS languages, where the CWO proportion is defined as the mean proportion of color terms that each participant used which were also used by more than three-quarters of the other participants. The non-normality of this distribution suggests that different tasks were used across different WCS languages: a free-choice version and a fixed-choice version. A proportion close to one suggests that the WCS task was implemented with a set of fixed choices for that language. Proportions much less than one suggest that the WCS task was implemented with free choice of color terms for that language. We include the Tsimane' fixed-choice and free-choice WCO proportions as baselines.

Instructions for the current study. We used two versions of a color-naming task: a free-choice version, in which participants were simply asked to label Munsell chips spanning the color space in a way that they thought others from their community would also label them; and a fixed-choice version, in which the instructions were identical to the free-choice version, but participants were also asked to choose from a fixed set of 8 choices (the modal labels from the free-choice version). In pilot experiments on 12 Tsimane' participants, we collected color-labeling data on the 160 chips of the standard Munsell array (6); subsequent participants were tested using a subset of 80 chips, sampling the array uniformly (the 80-chip array produced the same results as the 160-chip array, but took half the time for data collection on each participant).

We provide a list of the Munsell chip designations for the chips we used in the **Table S2**. Each color chip was affixed to a white cardboard square 2 inches on a side.

Participants were presented with the 80 chips in a different random order for each participant under controlled lighting conditions using a light box. Color-naming variability measured in studies that do not control for viewing conditions could arise because of variations in ambient light, adding noise to the naming task. The WCS used a stereotyped order for all chips, which may have also introduced systematic response biases. Using a random order for every participant avoids this possibility. The chips were about 1.5" square, mounted on a white card, and presented one at a time. The task was performed indoors for all three groups: at MIT for English participants, at the CBIDSI headquarters in San Borja, Bolivia, for Spanish participants, and in the village school houses for the Tsimane' participants. For the Tsimane' version of the task, the light box was powered by a car battery which we transported to the Tsimane' villages.

The complete instructions for the task were as follows:

In Tsimane':
Ma'je' tsun chij mo'in coty cororsi' in Tsimanesćan
Medyes qui tsun ma'je' paj qui jitica mi'in mo'in coror in oij ches carta in.
Jevaj jedye' buty tsun jidiyaja' oij coror.
(Fixed-choice version of the task: Mo'ya 8 in: Tsincus, jaibas, jäinäs, yushñus, shandyes, itsidyeisi, cafedyeisi, chocoratedyeisi, judyeya chames.
Dyim tyeva' juñis buty mi arajdye' coij mo' coror.)

In Spanish:
Queremos saber los nombres de los colores en Español. Así que queremos que nos digas los colores de estas cartas. Dinos como la gente llamaría estas cartas en Español.
(Fixed-choice version of the task: Hay 12 opciones: negro, blanco, rojo, azul, celeste, verde, morado, cafe, amarillo, anaranjado, rosa, gris.
Escoge el nombre del color mas cercano.)

In English:
We want to know the words for colors in English. So we want you to tell us the colors of these cards. Tell us what other English speakers would typically call these cards.
(Fixed-choice version of the task: There are 11 choices: black, white, red, green, blue, purple, brown, yellow, orange, pink, grey. Choose the closest color word.)

English participants' use of complex color terms. Out of 31 English participants in the free-choice version of the task, 24 sometimes used multi-word color descriptors, such as "dark green" or "baby blue", resulting in 17.8% (436 / 2440) trials with multi-word color descriptors. We entered the head noun as the color for these descriptors (e.g., "dark green" was coded as "green"; "baby blue" as "blue"). Interestingly, the Bolivian-Spanish and Tsimane' participants never used multi-word color descriptors: they always used single word colors. The difference between English on the one hand and Spanish and Tsimane' on the other may partially arise from the pragmatics of the situation. The English speakers knew that the testers were native English speakers, and therefore the task became to label the colors as narrowly as possible (ignoring the instructions, such that participants are supposed to label colors as other English speakers in their community would). For the Tsimane' and Bolivian Spanish speakers, the task instructions were plausibly followed more closely, perhaps because the participants knew that the testers (E.G., M.G., J.J.-E.) were not native speakers of Tsimane' or Bolivian Spanish.

Consistent behavior of participants. All participants, in both versions of the task, showed above-chance categorization of the color chips into a color-partition space, thus ensuring that our results could not be explained by poor color detection in some groups or participants (**Figures S3-S5** show sample color response grids for 5 randomly chosen speakers from each of the three languages).

To ensure that our results could not be explained by participants randomly assigning color words to color chips, we confirmed that each participant was responding to the task in a consistent way. To do this, we tested if the number of color word clusters generated by each participant was significantly smaller than expected if the participant were selecting color words from their vocabulary at random. To do so we first defined a cluster as a group of adjacent chips (horizontally, vertically, or diagonally) for which the speaker had chosen the same color word. After computing the number of color word clusters that each participant produced in the task, we calculated the probability of observing a number of clusters as low as the true number through a permutation test with 100 samples. That is, for each participant we generated a baseline distribution by randomly rearranging the color words 100 times and calculating the resulting number of clusters each time. By comparing these 100 baseline clusters with the true number of clusters that each participant produced, it is possible to determine the likelihood that participants were simply uttering color words at random. Critically, this analysis is both sensitive to the number of color words each participant used, and to the frequency with which they used each word. On average, participants produced 17 color-word clusters. In contrast, the average baseline number of clusters expected by chance was 46. Moreover, for all participants in all languages (English, Spanish, and Tsimane') and both tasks (fixed-choice and free-choice versions), all baseline samples produced a strictly larger number of clusters than the ones participants produced. The probability that participants could have produced such a structured division of the grid space by chance is $p < 0.001$.

Our Code	Munsell Code	WCS Code	In labeling experiment?	In focal color experiment?	In the 24 chips evenly sampling CIELAB?	In RT experiment?
A1	5R9/2	B1	FALSE	TRUE	FALSE	FALSE
A2	10R9/2	B3	TRUE	TRUE	FALSE	FALSE
A3	5YR9/2	B5	FALSE	TRUE	FALSE	FALSE
A4	10YR9/4	B7	TRUE	TRUE	FALSE	FALSE
A5	5Y9/6	B9	FALSE	TRUE	FALSE	FALSE
A6	10Y9/6	B11	TRUE	TRUE	TRUE	FALSE
A7	5GY9/4	B13	FALSE	TRUE	FALSE	FALSE
A8	10GY9/4	B15	TRUE	TRUE	TRUE	FALSE
A9	5G9/2	B17	FALSE	TRUE	FALSE	FALSE
A10	10G9/2	B19	TRUE	TRUE	FALSE	FALSE
A11	5BG9/2	B21	FALSE	TRUE	FALSE	FALSE
A12	10BG9/2	B23	TRUE	TRUE	FALSE	FALSE
A13	5B9/2	B25	FALSE	TRUE	FALSE	FALSE
A14	10B9/2	B27	TRUE	TRUE	FALSE	FALSE
A15	5PB9/2	B29	FALSE	TRUE	FALSE	FALSE
A16	10PB9/2	B31	TRUE	TRUE	FALSE	TRUE
A17	5P9/2	B33	FALSE	TRUE	FALSE	FALSE
A18	10P9/2	B35	TRUE	TRUE	FALSE	FALSE
A19	5RP9/2	B37	FALSE	TRUE	FALSE	FALSE
A20	10RP9/2	B39	TRUE	TRUE	FALSE	FALSE
B1	5R8/6	C1	TRUE	TRUE	FALSE	TRUE
B2	10R8/6	C3	FALSE	TRUE	FALSE	FALSE
B3	5YR8/8	B5	TRUE	TRUE	FALSE	FALSE
B4	10YR8/14	C7	FALSE	TRUE	FALSE	TRUE
B5	5Y8/14	C9	TRUE	TRUE	FALSE	FALSE
B6	10Y8/12	C11	FALSE	TRUE	FALSE	FALSE
B7	5GY8/10	C13	TRUE	TRUE	FALSE	FALSE
B8	10GY8/8	C15	FALSE	TRUE	FALSE	FALSE
B9	5G8/6	C17	TRUE	TRUE	TRUE	FALSE
B10	10G8/6	C19	FALSE	TRUE	FALSE	FALSE
B11	5BG8/4	C21	TRUE	TRUE	FALSE	FALSE
B12	10BG8/4	C23	FALSE	TRUE	FALSE	FALSE
B13	5B8/4	C25	TRUE	TRUE	FALSE	TRUE
B14	10B8/6	C27	FALSE	TRUE	FALSE	FALSE
B15	5PB8/6	C29	TRUE	TRUE	FALSE	FALSE
B16	10PB8/4	C31	FALSE	TRUE	FALSE	FALSE
B17	5P8/4	C33	TRUE	TRUE	FALSE	FALSE
B18	10P8/6	C35	FALSE	TRUE	FALSE	FALSE
B19	5RP8/6	C37	TRUE	TRUE	FALSE	FALSE
B20	10RP8/6	C39	FALSE	TRUE	FALSE	FALSE
C1	5R7/10	D1	FALSE	TRUE	FALSE	FALSE
C2	10R7/10	D3	TRUE	TRUE	TRUE	FALSE

C3	5YR7/14	D5	FALSE	TRUE	FALSE	FALSE
C4	10YR7/14	D7	TRUE	TRUE	FALSE	FALSE
C5	5Y7/12	D9	FALSE	TRUE	FALSE	FALSE
C6	10Y7/12	D11	TRUE	TRUE	FALSE	FALSE
C7	5GY7/12	D13	FALSE	TRUE	FALSE	FALSE
C8	10GY7/10	D15	TRUE	TRUE	FALSE	FALSE
C9	5G7/10	D17	FALSE	TRUE	FALSE	FALSE
C10	10G7/8	D19	TRUE	TRUE	TRUE	FALSE
C11	5BG7/8	D21	FALSE	TRUE	FALSE	FALSE
C12	10BG7/8	D23	TRUE	TRUE	FALSE	FALSE
C13	5B7/8	D25	FALSE	TRUE	FALSE	FALSE
C14	10B7/8	D27	TRUE	TRUE	FALSE	FALSE
C15	5PB7/8	D29	FALSE	TRUE	FALSE	FALSE
C16	10PB7/8	D31	TRUE	TRUE	FALSE	FALSE
C17	5P7/8	D33	FALSE	TRUE	FALSE	FALSE
C18	10P7/8	D35	TRUE	TRUE	FALSE	FALSE
C19	5RP7/10	D37	FALSE	TRUE	FALSE	FALSE
C20	10RP7/8	D39	TRUE	TRUE	FALSE	FALSE
D1	5R6/12	E1	TRUE	TRUE	FALSE	FALSE
D2	10R6/14	E3	FALSE	TRUE	FALSE	FALSE
D3	5YR6/14	E5	TRUE	TRUE	FALSE	FALSE
D4	10YR6/12	E7	FALSE	TRUE	FALSE	FALSE
D5	5Y6/10	E9	TRUE	TRUE	FALSE	TRUE
D6	10Y6/10	E11	FALSE	TRUE	FALSE	FALSE
D7	5GY6/10	E13	TRUE	TRUE	FALSE	FALSE
D8	10GY6/12	E15	FALSE	TRUE	FALSE	FALSE
D9	5G6/10	E17	TRUE	TRUE	FALSE	FALSE
D10	10G6/10	E19	FALSE	TRUE	FALSE	FALSE
D11	5BG6/10	E21	TRUE	TRUE	FALSE	FALSE
D12	10BG6/8	E23	FALSE	TRUE	FALSE	FALSE
D13	5B6/10	E25	TRUE	TRUE	FALSE	FALSE
D14	10B6/10	E27	FALSE	TRUE	FALSE	FALSE
D15	5PB6/10	E29	TRUE	TRUE	FALSE	FALSE
D16	10PB6/10	E31	FALSE	TRUE	FALSE	FALSE
D17	5P6/8	E33	TRUE	TRUE	FALSE	FALSE
D18	10P6/10	E35	FALSE	TRUE	FALSE	FALSE
D19	5RP6/12	E37	TRUE	TRUE	TRUE	FALSE
D20	10RP6/12	E39	FALSE	TRUE	FALSE	FALSE
E1	5R5/14	F1	FALSE	TRUE	FALSE	FALSE
E2	10R5/16	F3	TRUE	TRUE	FALSE	FALSE
E3	5YR5/12	F5	FALSE	TRUE	FALSE	FALSE

E4	10YR5/10	F7	TRUE	TRUE	FALSE	FALSE
E5	5Y5/8	F9	FALSE	TRUE	FALSE	FALSE
E6	10Y5/8	F11	TRUE	TRUE	FALSE	FALSE
E7	5GY5/10	F13	FALSE	TRUE	FALSE	FALSE
E8	10GY5/12	F15	TRUE	TRUE	FALSE	FALSE
E9	5G5/10	F17	FALSE	TRUE	FALSE	FALSE
E10	10G5/10	F19	TRUE	TRUE	FALSE	FALSE
E11	5BG5/10	F21	FALSE	TRUE	FALSE	FALSE
E12	10BG5/10	F23	TRUE	TRUE	TRUE	FALSE
E13	5B5/10	F25	FALSE	TRUE	FALSE	FALSE
E14	10B5/12	F27	TRUE	TRUE	TRUE	FALSE
E15	5PB5/12	F29	FALSE	TRUE	FALSE	FALSE
E16	10PB5/10	F31	TRUE	TRUE	TRUE	FALSE
E17	5P5/10	F33	FALSE	TRUE	FALSE	FALSE
E18	10P5/12	F35	TRUE	TRUE	TRUE	FALSE
E19	5RP5/12	F37	FALSE	TRUE	FALSE	FALSE
E20	10RP5/14	F39	TRUE	TRUE	FALSE	FALSE
F1	5R4/14	G1	TRUE	TRUE	FALSE	TRUE
F2	10R4/12	G3	FALSE	TRUE	FALSE	FALSE
F3	5YR4/8	G5	TRUE	TRUE	TRUE	FALSE
F4	10YR4/8	G7	FALSE	TRUE	FALSE	FALSE
F5	5Y4/6	G9	TRUE	TRUE	TRUE	TRUE
F6	10Y4/6	G11	FALSE	TRUE	FALSE	FALSE
F7	5GY4/8	G13	TRUE	TRUE	TRUE	FALSE
F8	10GY4/8	G15	FALSE	TRUE	FALSE	FALSE
F9	5G4/10	G17	TRUE	TRUE	FALSE	TRUE
F10	10G4/10	G19	FALSE	TRUE	FALSE	FALSE
F11	5BG4/8	G21	TRUE	TRUE	TRUE	FALSE
F12	10BG4/8	G23	FALSE	TRUE	FALSE	FALSE
F13	5B4/10	G25	TRUE	TRUE	TRUE	TRUE
F14	10B4/10	G27	FALSE	TRUE	FALSE	FALSE
F15	5PB4/12	G29	TRUE	TRUE	TRUE	FALSE
F16	10PB4/12	G31	FALSE	TRUE	FALSE	FALSE
F17	5P4/12	G33	TRUE	TRUE	TRUE	TRUE
F18	10P4/12	G35	FALSE	TRUE	FALSE	FALSE
F19	5RP4/12	G37	TRUE	TRUE	TRUE	FALSE
F20	10RP4/14	G39	FALSE	TRUE	FALSE	FALSE
G1	5R3/10	H1	FALSE	TRUE	FALSE	FALSE
G2	10R3/10	H3	TRUE	TRUE	TRUE	FALSE
G3	5YR3/6	H5	FALSE	TRUE	FALSE	FALSE
G4	10YR3/6	H7	TRUE	TRUE	FALSE	TRUE

G5	5Y3/4	H9	FALSE	TRUE	FALSE	FALSE
G6	10Y3/4	H11	TRUE	TRUE	FALSE	FALSE
G7	5GY3/6	H13	FALSE	TRUE	FALSE	FALSE
G8	10GY3/6	H15	TRUE	TRUE	TRUE	FALSE
G9	5G3/8	H17	FALSE	TRUE	FALSE	FALSE
G10	10G3/8	H19	TRUE	TRUE	FALSE	FALSE
G11	5BG3/8	H21	FALSE	TRUE	FALSE	FALSE
G12	10BG3/8	H23	TRUE	TRUE	FALSE	FALSE
G13	5B3/8	H25	FALSE	TRUE	FALSE	FALSE
G14	10B3/10	H27	TRUE	TRUE	TRUE	TRUE
G15	5PB3/10	H29	FALSE	TRUE	FALSE	FALSE
G16	10PB3/10	H31	TRUE	TRUE	TRUE	FALSE
G17	5P3/10	H33	FALSE	TRUE	FALSE	FALSE
G18	10P3/10	H35	TRUE	TRUE	FALSE	FALSE
G19	5RP3/10	H37	FALSE	TRUE	FALSE	FALSE
G20	10RP3/10	H39	TRUE	TRUE	FALSE	FALSE
H1	5R2/8	I1	TRUE	TRUE	TRUE	FALSE
H2	10R2/6	I3	FALSE	TRUE	FALSE	FALSE
H3	5YR2/4	I5	TRUE	TRUE	FALSE	FALSE
H4	10YR2/2	I7	FALSE	TRUE	FALSE	FALSE
H5	5Y2/2	I9	TRUE	TRUE	FALSE	FALSE
H6	10Y2/2	I11	FALSE	TRUE	FALSE	FALSE
H7	5GY2/2	I13	TRUE	TRUE	FALSE	TRUE
H8	10GY2/4	I15	FALSE	TRUE	FALSE	FALSE
H9	5G2/6	I17	TRUE	TRUE	FALSE	TRUE
H10	10G2/6	I19	FALSE	TRUE	FALSE	FALSE
H11	5BG2/6	I21	TRUE	TRUE	FALSE	FALSE
H12	10BG2/6	I23	FALSE	TRUE	FALSE	FALSE
H13	5B2/6	I25	TRUE	TRUE	FALSE	FALSE
H14	10B2/6	I27	FALSE	TRUE	FALSE	FALSE
H15	5PB2/8	I29	TRUE	TRUE	TRUE	FALSE
H16	10PB2/10	I31	FALSE	TRUE	FALSE	FALSE
H17	5P2/8	I33	TRUE	TRUE	FALSE	TRUE
H18	10P2/6	I35	FALSE	TRUE	FALSE	FALSE
H19	5RP2/8	I37	TRUE	TRUE	FALSE	FALSE
H20	10RP2/8	I39	FALSE	TRUE	FALSE	FALSE

Table S2. The 160 Munsell chips that were used in our experiments. As indicated in the rightmost four columns, 80 of these color chips were used in the labeling experiment; all 160 were used in the focal color determination; 24 were used in the analysis of CIELAB colors; and 15 were used in the reaction time (RT) experiment.

Data from individual subjects. Here, we show the responses of 5 randomly chosen speakers from each of the three languages for the Munsell-chip free-choice color-naming experiment. Each color word is given a unique color, and the color of the chip for a given speaker reflects the color word used for that chip by that speaker. The colors used for the main color words in English and Bolivian Spanish are assigned based on the focal colors for those words. For Tsimane', we take the modal focal color (mode focal hue, mode focal luminance) for each color.

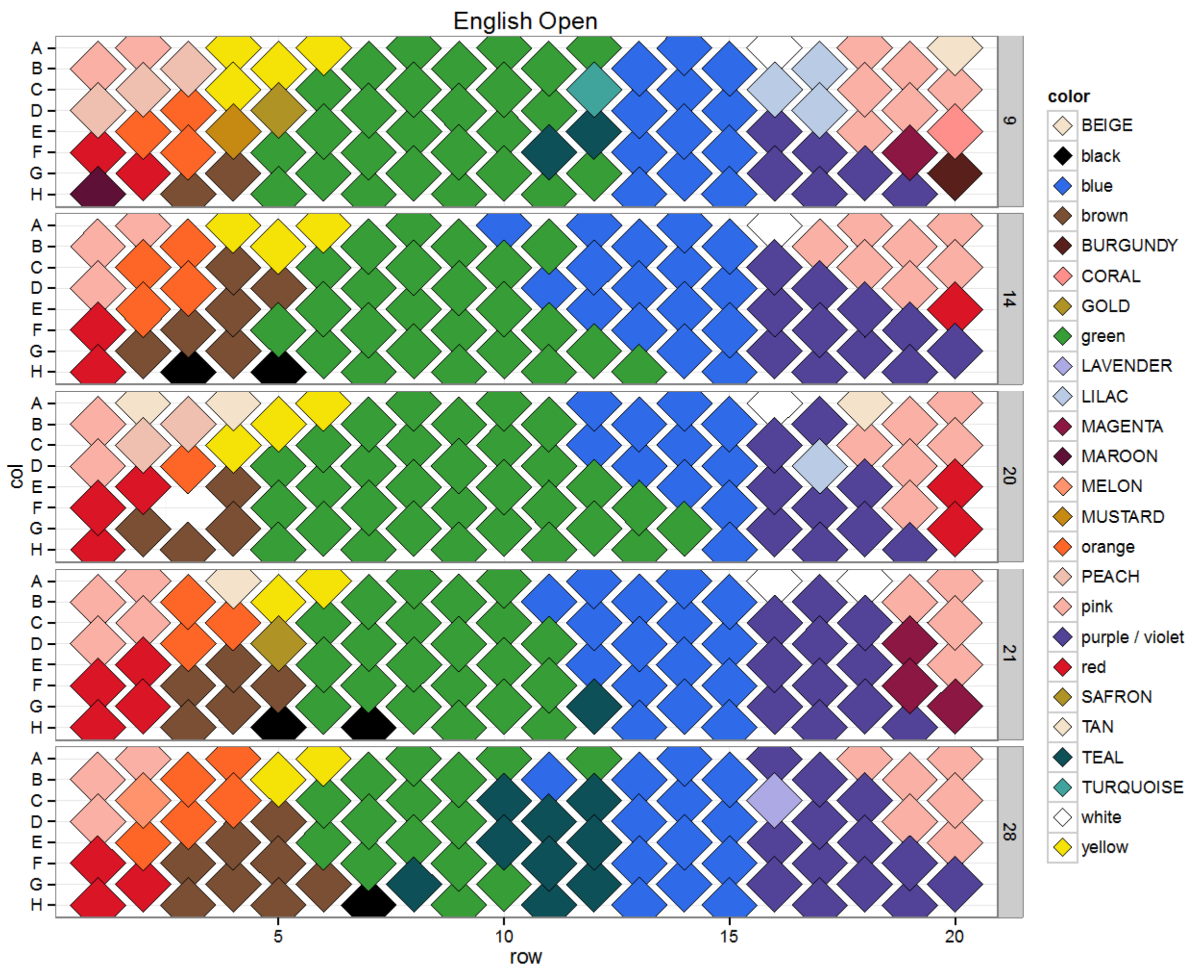

Figure S3. Sample color grids for 5 randomly chosen speakers from English using the free-choice paradigm, in which participants could label the colors without any restrictions on the labels they could use.

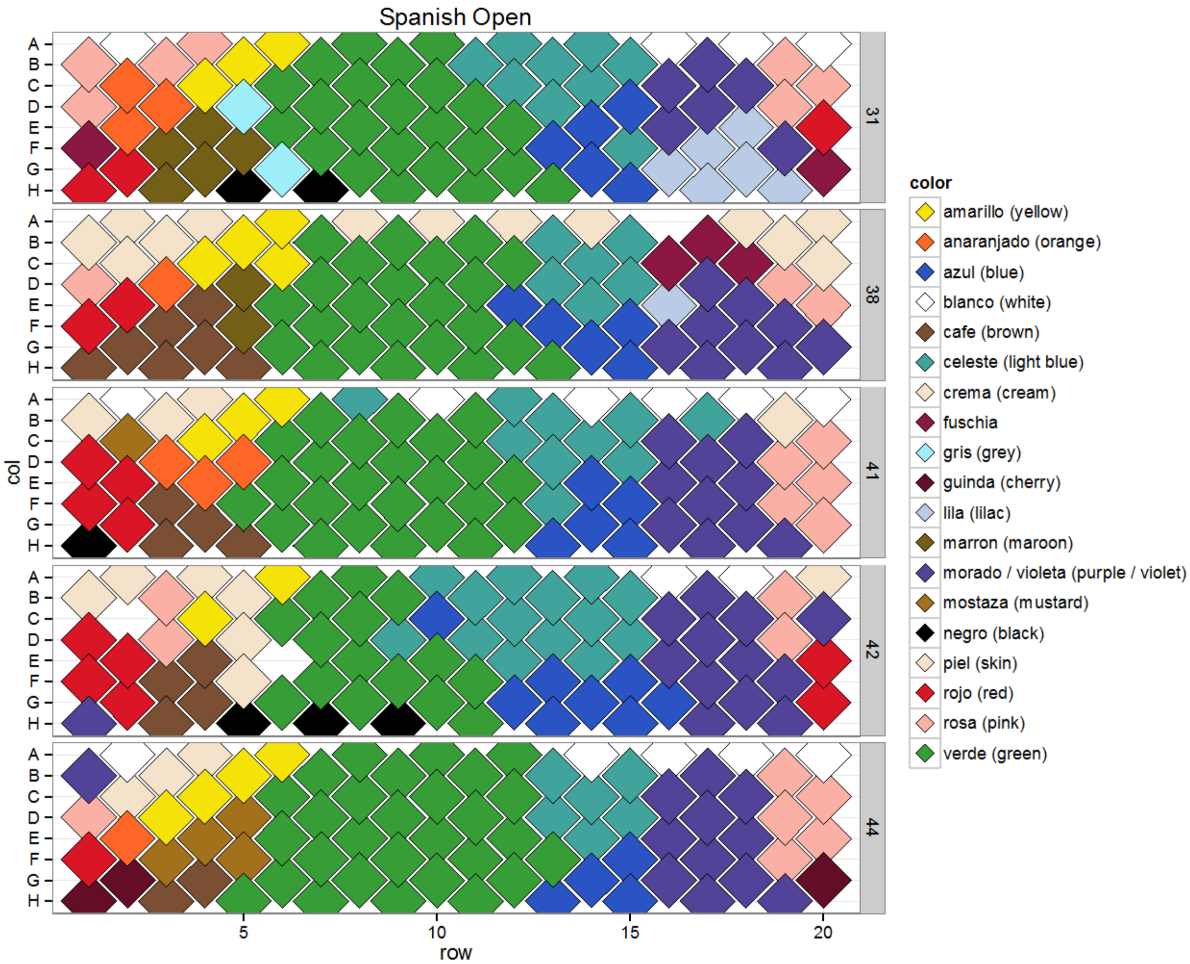

Figure S4. Sample color grids for 5 randomly chosen speakers from Bolivian Spanish using the free-choice paradigm, in which participants could label the colors without any restrictions on the labels they could use.

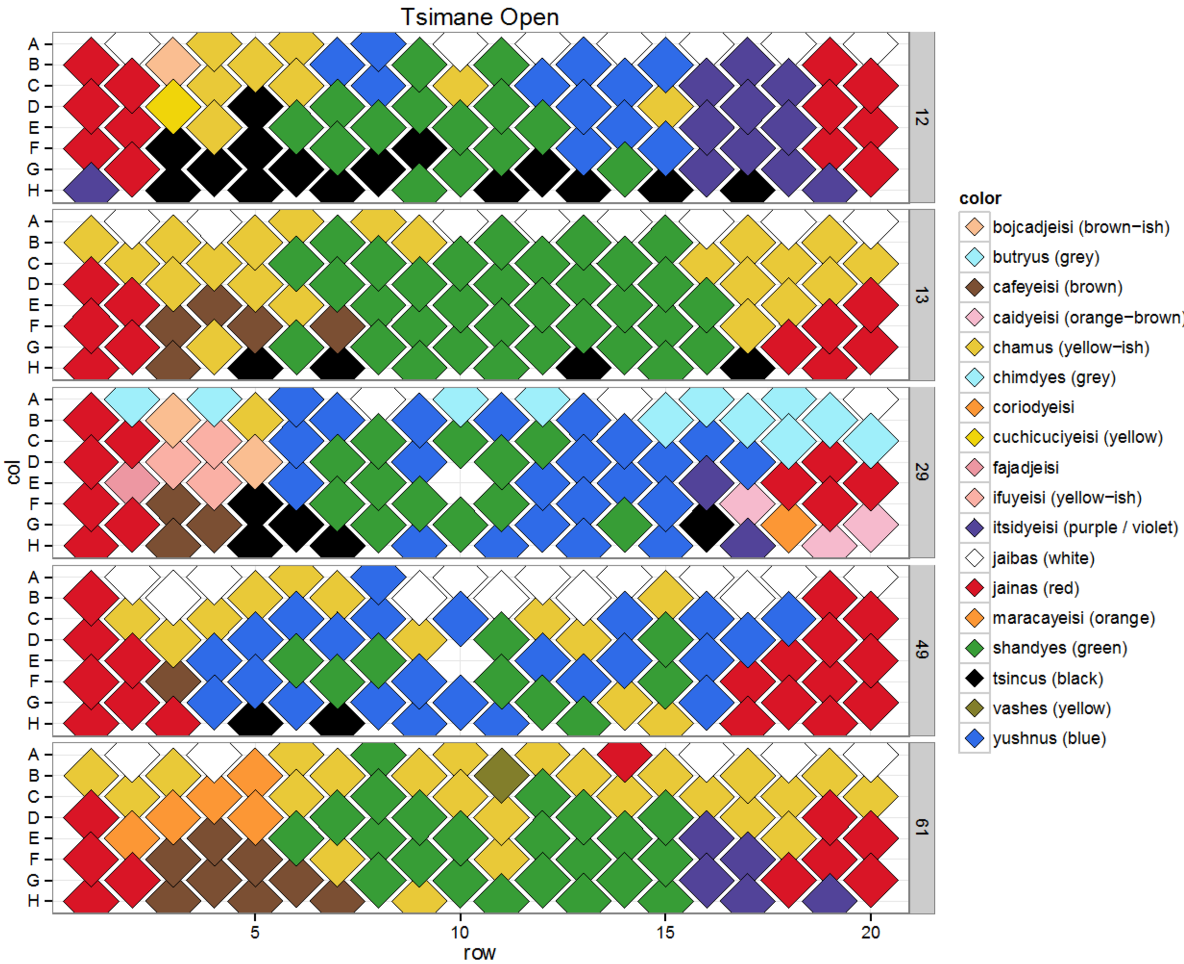

Figure S5. Sample color grids for 5 randomly chosen speakers from Tsimane' using the free-choice paradigm, in which participants could label the colors without any restrictions on the labels they could use.

Color in Fig. 1	Spanish	English	Tsimane'
	blanco (100%, 100%)	white (100%, 100%)	*jaibas* (100%, 100%)
	negro (100%, 100%)	black (100%, 100%)	*tsincus* (100%, 100%)
	rojo (100%, 95%)	red (100%, 97%)	*jäinäs* (100%, 100%)
	verde (100%, 100%)	green (100%, 100%)	*shandyes* (91%, 62%)
	amarillo (100%, 95%)	yellow (100%, 97%)	*chames* (79%, 43%)
	-	blue (100%, 97%)	*yụshñus* (78%, 57%)
	marrón (95%, 85%)	brown (100%, 100%)	*cafedyeisi / chocoratedyeisi* (74%, 52%)
	púrpura (95%, 85%)	purple (97%, 100%)	*itsidyeisi* (64%, 40%)
	naranja (100%, 85%)	orange (97%, 87%)	-
	rosado (95%, 95%)	pink (100%, 100%)	-
	celeste (100%, 95%)	-	-
	azul (100%, 100%)	-	-

Table S3. Empirically determined "Basic Color Terms" in Bolivian Spanish, English and Tsimane'. The first percentage is the fraction of each population that used the term at least once in naming any color in the 80-chip free-choice color-naming task; the second percentage is the largest modal value for that color term among all the color chips in the free-choice task Corresponding terms across languages are identified using data from **Figure 1**. Terms have been rank-ordered top-to-bottom according to frequency of use in Tsimane'. The color in the left column provides a key with the results in **Figure 1**. Note that the word for "color" in Tsimane' is "yeisi" (often shortened to "yes / -s"). All of the color words that we encountered are native (non-borrowed) Tsimane' except the word for brown: *cafedyeisi / chocoratedyeisi*, borrowed from Spanish.

The average surprisal analysis results of the fixed-options version of the task were strikingly similar to those from the free-choice response task (compare **Figure 1** with **Figure S6**). The average surprisal of each language hardly changes at all from one task to the other: Tsimane': 4.88 bits in free-choice; 4.91 in fixed-choice; English: 3.80 bits in free-choice; 3.86 in fixed-choice; Bolivian Spanish: 3.86 bits in free-choice; 3.94 in fixed-choice. This demonstrates that the free-choice and the fixed-choice tasks (the second of which is more similar to the WCS task) provide strikingly similar results, suggesting a robustness of results to particular testing procedures for color labeling tasks.

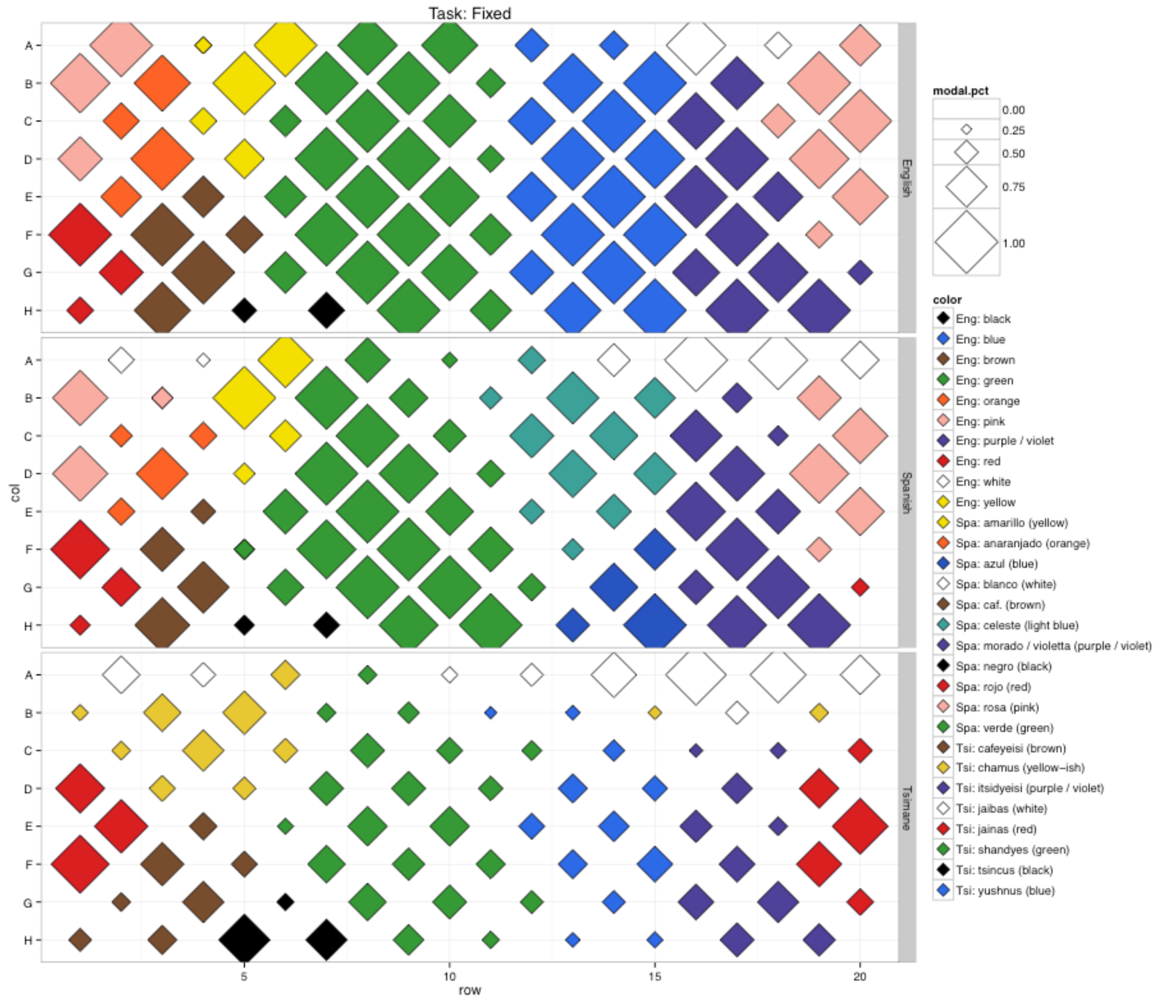

Figure S6. Diamond plots of the population responses for English, Spanish and Tsimane' in the color-labeling task where participants had a fixed set of possible choices. Each chip that was presented to the participant is shown using the modal color word used for that chip, where each color word is represented by a different color. The diameter of the diamond is the proportion of participants that use the modal color word for that chip (Similar conclusions were obtained using the free-choice version of the task; compare with **Figure 1** in the main text).

SI-Section 2: Control experiment with Tsimane' and English speakers: Reaction times for naming objects and colors

We performed a control experiment to ensure that the participants were fully engaged in the various tasks. We assessed the time required to label 15 colored chips spanning the Munsell array (including focal and boundary colors; **Table S2**), and eight common Tsimane' objects (a ripe banana, a ripe tomato, a rock, a stick, a leaf, a comb, a cup, and a fan (Tsimane' artifact)), which were physically presented to each participant (**Figure S7**).

Each participant received a different random order of the objects and colors. The participants consisted of 66 Tsimane' adults (mean age: 31.4 years; SD: 14.2 years; range 17-85; 44 females) recruited from 3 Tsimane' communities near San Borja, Bolivia, and 23 English participants (mean age: 26.5 years; SD: 10.9 years; range 18-58; 10 females) recruited from the local MIT community. We video-recorded all trials. Two coders independently timed each of the English and Tsimane' videos. We used the average time of these measurements in our analyses, analyzing over all trials

We fit a mixed effect linear regression predicting log color chip naming latency time from language and the entropy of the color chip, as defined in equation (2) in the main article. We included random intercepts for participant and color with a random slope by language for the object label. We normalized the entropy predictor. We found that increased entropy led to significantly higher naming latency in log seconds (beta=.25, t = 7.515, $p < .0001$). There was also a main effect for English reaction times to be faster compared to Tsimane' reaction times (beta=-.19, t=-3.93, $p < .0001$). There was also no significant interaction although there was a trend for the slope of entropy to be less steep in English (beta=-.06, chisq(1) = 3.40, p = .07).

For object naming latencies, there was again a main effect of entropy on latency (beta=.22, t=4.74, $p < .0001$). There was no clear effect of language, and if anything English was slower for object naming than Tsimane' (beta=.13, t=1.81, chi2(1) = 3.39, p = .07) by a chi-squared likelihood ratio test). There was a trend for the entropy effect to be greater for English (beta=.10 t=1.84, chisq(1) = 3.57, p =.06) although that trend is largely driven by the large average RT for the object "stick" (which received many labels and elicited long latencies in English but not Tsimane') and we should therefore not conclude much from it.

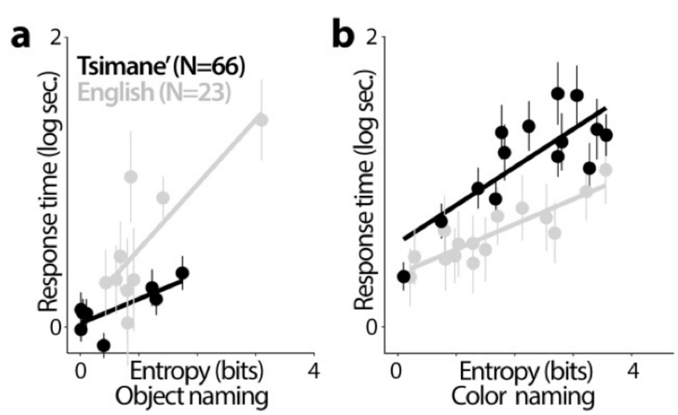

Figure S7. Reaction time to naming objects (**a**, 8 objects) and colors (**b**, 15 colors) as a function of the entropy of each object or color. Increased entropy correlated with higher latency. For objects: beta=.22, t=4.86, $p < .0001$; English tended to be slower, although insignificantly (beta=.12, t=1.66, chi2(1) = 2.90, p = .09). For colors: beta=.25, t = 7.34, $p < .0001$; main effect for English reaction times to be faster compared to Tsimane' reaction times (beta=-.19, t=-4.08, $p < .0001$). Error bars show 95% confidence intervals on the mean log reaction time for each chip or object.

SI-Section 3: Computing average surprisal for each chip

The average surprisal scores for each chip, in the three languages, is given in **Figure S8**.

By equation 1, the average surprisal score for a color chip c is:

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

For example, suppose a particular color chip is labeled with four different words across the population, in the following distribution:

C1: W_1: 50%; W_2: 30%; W_3: 15%; W_4: 5%

these are the *P(w|c)*: the probabilities that a particular color c gets labeled as w

We also need the surprisal for each color word: *-log P(c|w)*. We can compute the *P(c|w)* by Bayes theorem:

*= P(w|c) * P(c) / P(w)*

We assume *P(c)* is uniform over the color space (= 1/80 for our 80 color chips), and we can compute *P(w)* across the color space: how often a particular word gets used, across participants. Suppose in this example that w has the following uses across the color space (suggesting equal use across the color space):

W_1: 20%; W_2: 20%; W_3: 20%; W_4: 20%
listener surprisals for W_1- W_4: for each W, *P(w|c) * P(c) / P(w)*
W_1: -log (.5 * 1/80 / .2) = 5
W_2: -log (.3 * 1/80 / .2) = 5.737
W_3: -log (.15 * 1/80 / .2) = 6.737
W_4: -log (.05 * 1/80 / .2) = 8.322

S(C1) = (.5 * 5) + (.3 * 5.737) + (.15 * 6.737) + (.05 * 8.322) = (2 + 1.72 + 1.01 + .416) = 5.65

This means that it would take about 5.65 bits of information to transfer this particular color to a listener. This is a lot of yes-no-questions because there aren't very many color words in this particular example vocabulary (four of the words are 80% of the words that people say), and there are a lot of colors to transmit (80).

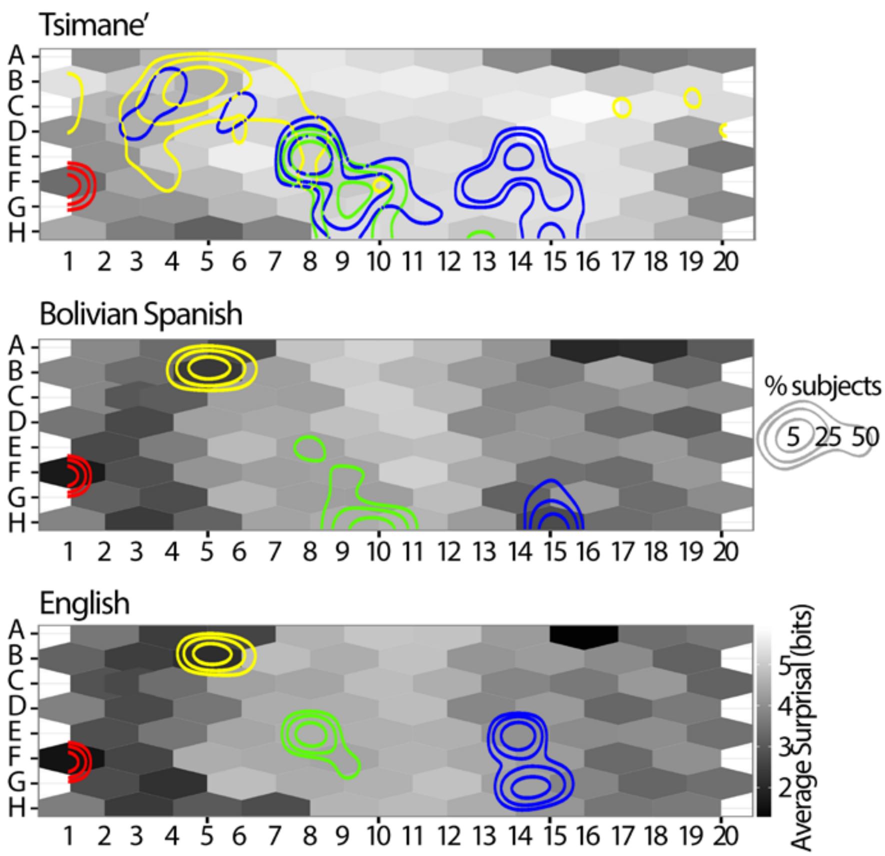

Figure S8. Average surprisal for each chip in the Munsell array, computed using equation 1, using data obtained in the free-choice version of the task. Data from the fixed-choice version of the task yielded similar results. The pattern of average surprisal across the three languages is similar, even if the overall average surprisal across the languages differs. Overlay shows independent data that captures the probability density of color samples chosen as the best examples for color words *rojo*, *verde*, *azul*, *amarillo* (Bolivian-Spanish, N=55); *red, green, blue, yellow* (English, N=29); and *jäinäs, shandyes, yụshñus, chamus* (Tsimane’, N=99). The contours enclose 5%, 25%, and 50% of the data. The four colors are the “unique hues”, which might have been predicted to show relatively low average surprisal. Instead, only the yellow and red chips showed high surprisal in all three languages.

SI-Section 4: Analyses of average surprisal within the World Color Survey data

In order to compare our findings with the WCS we computed the informativity of each language for the common 80 chips and we compared it with the number of color words used. **Figure S9** shows the relation between number of color words and the average surprisal across languages (see **Figure 3A**). As expected, languages with more color terms tend to have less uncertainty. Spanish and English show the lowest uncertainty compared to other languages with a similar number of color words. Estimates of average surprisal across the WCS uncovered a broad diversity of color-systems among the world’s languages (**Figure S9**, smaller open circles); Tsimane’ is representative of most color systems in the WCS. In addition, as the average number of words increases across the population of languages, the average surprisal of the languages decreases: in general, languages with more color terms have more informative color systems.

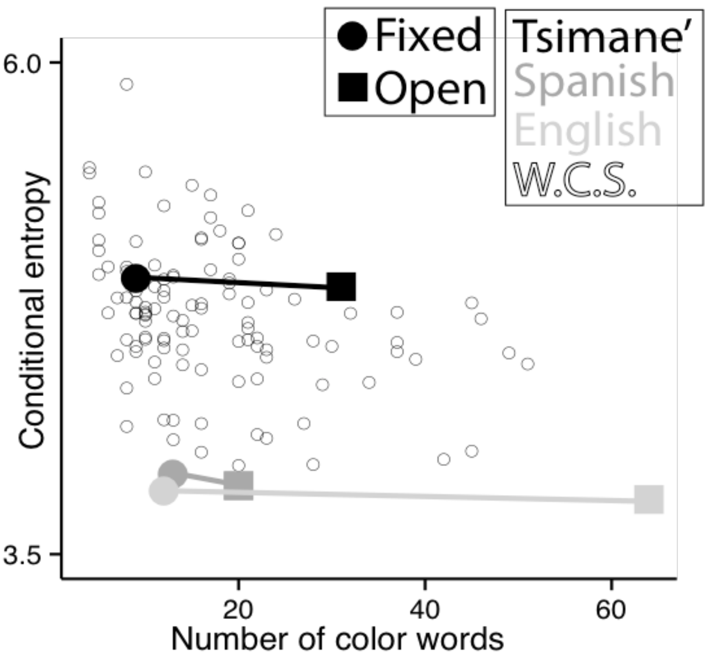

Figure S9. Average surprisal within a language versus the total number of color words used in each population. World Color Survey (small open circles) and the three populations tested here (solid symbols). Circles show data from experiments in which participants were constrained to use a fixed vocabulary of basic color terms; squares show data where participants were free to use any term. The average surprisal is similar for the free-choice and fixed-choice versions of the task in English, Spanish and Tsimane'. English and Spanish have lower average surprisal values than the languages in the WCS (the WCS comprises predominantly non-industrialized cultures; data replotted from **Figure 3A**).

To ensure the validity of our results we repeated the same analysis after filtering uncommon words in all languages. To do so, we filtered out all color words for which the percentage of participants using these words did not surpass thresholds of 20% and 50%, as shown in **Figures S10** and **S11**. Critically, the average surprisal values remain roughly constant for the free and fixed-choice versions of the task in Tsimane', English and Spanish, for the 0, 20 and 50% thresholds, showing the robustness of the task and results.

Our results suggest that the most robust complexity metric to use when comparing color-naming across languages is a trial-based measure of information, such as average surprisal (equation 2) or mutual information (Lindsey et al, 2015) rather than the number of (basic) color words that the language uses (Berlin & Kay, 1969). In particular, average surprisal provides a consistent measure across different versions of the color-naming task, and it provides a trial-based measure which takes into account the consistency of labeling a particular color across participants. Interestingly, Tsimane' turns out to have a less sophisticated color-naming system than the bulk of the world's languages: we can see from **Figure 4** that 82 of the 110 WCS languages have more information in their color-naming systems than Tsimane'. This is the case in spite of the fact that Tsimane' has 8 modal color names across its color grid, more than many languages which have more information in their color-naming systems than Tsimane' has. This is because Tsimane' has relatively low agreement across participants on what to call each color. In particular, in the free-choice version of the color-naming task, 46 of the 80 color chips that participants labeled had modal labels of below 50%. This contrasts with Pirahã from the WCS, for example, which had only four modal color words (in a fixed-choice labeling paradigm), but where participants had much higher agreement on each color chip. Under an information-theoretic analysis, Tsimane' and Pirahã transmit similar amounts of information with their labeling systems.

Situating Tsimane' in Berlin & Kay's proposed color-word complexity space is difficult. There are several irregularities. For example, it might seem that the word *chames* corresponds roughly to "yellow" in Berlin & Kay's ordered color hierarchy, and that it might enter the language fifth, by the percentages in **Table S3**, such that 79% of participants use this color word. Upon closer inspection however, one sees that *chames* is not used regularly by participants, in spite of the fact that most people know the word. Indeed, although there were 8 color chips for which the modal label was *chames*, these modal values were very low: between 17% and 43%. So while *chames* is a color word that many participants use, it does not have a standardized meaning within the language yet.

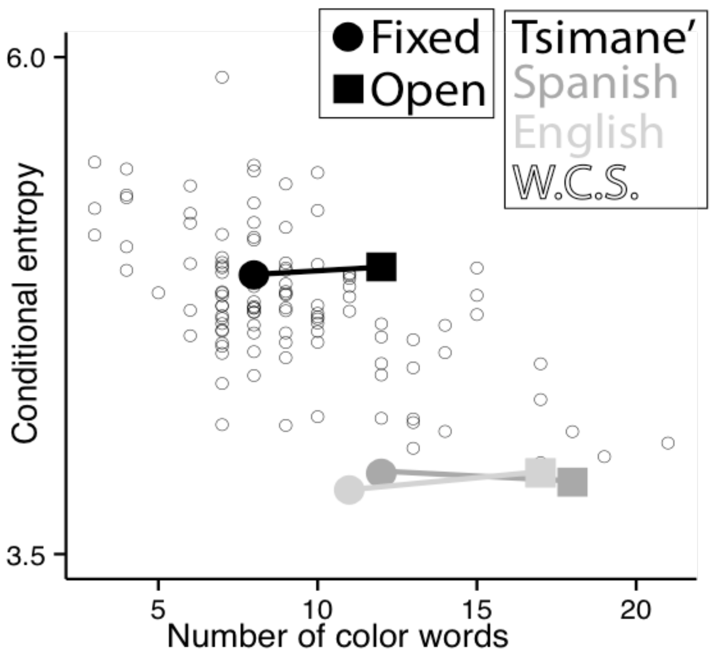

Figure S10. Average surprisal within a language versus the number of color words, filtered to only those color words that were provided by at least 20% of participants. The average surprisal values for the free and fixed-choice versions of the task remain roughly constant in English, Spanish and Tsimane' as in **Figure S9** (compare with **Figure 3A**).

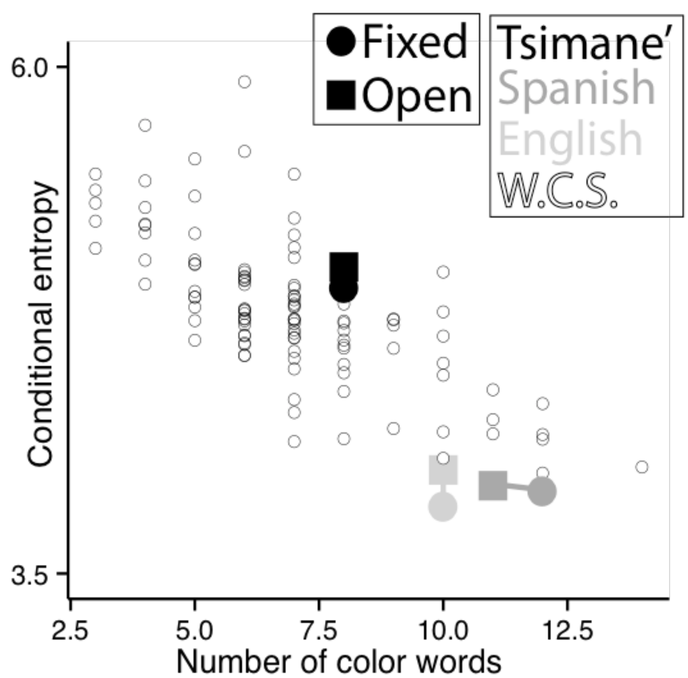

Figure S11. Average surprisal within a language versus the number of color words, filtered to only those color words that were provided by at least 50% of participants. The average surprisal values for the free and fixed-choice versions of the task remain roughly constant in English, Spanish and Tsimane' as in **Figure S9** (compare with **Figure 3A**).

SI-Section 5: Focal colors & unique hues

Following the Munsell-chip color naming experiment, each participant (N=99 Tsimane'; 55 Spanish; 29 English) was then presented with a standard 160-chip Munsell array of colors (illuminated by the lightbox), and was asked to point out the best example of several color words. The array of colors was organized by a 8 x 20 grid, mounted on matte black cardboard, and each color was a square about 0.5cm across, separated from other colored squares by ~3mm. We indexed the colors A-H according to lightness, and 1-20 according to hue. The chips most often selected as focal colors for all the terms probed are given in **Table S4**. To show the population results and evaluate the possible privilege of the unique hues, we computed the probability density function for each of the four unique hues over the grid space. The contours in **Figure S8** show the probability that a given color word was used for each color chip, on the basis of our empirical data. The lines show boundaries inside which probability mass is 5%, 25%, and 50%. The probability density functions were obtained through cubic spline interpolation on the color grid. The probability density functions were computed in Python using the "zoom" function in the *scipy* package, and the contours were calculated using the *matplotlib* package. The rank-ordering of the colors by communication efficiency was not predicted by the unique hues (**Table S5)**.

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

Table S4. Most frequently chosen chips as best examples of the color terms queried,

Rank	English	Spanish	Tsimane'
1	A16	F1	H5
2	F1	A16	A16
3	B5	A18	F1
4	G4	B5	A18
5	B3	D3	E20
6	H3	H3	A20
7	D3	E2	H3
8	A4	A6	H7
9	E2	H15	G20
10	A6	G4	A14
11	F3	F3	A2
12	C2	C4	G2
13	G2	A20	E2
14	H7	D19	D1
15	H5	H5	G4
16	B19	G2	F19
17	B1	G14	D19
18	C20	B3	H1
19	G18	B19	D3
20	D17	A4	F3
21	C4	F17	B5
22	F17	G18	C4
23	A18	H17	H19
24	G20	D17	C2
25	E20	F15	G6
26	E4	D1	A4
27	D5	H19	F5
28	H17	E20	B3
29	A2	E4	G18
30	A20	C2	B17
31	C16	B15	H13
32	H19	H1	D5
33	D19	B1	A6
34	H1	C14	F9
35	D1	A2	E4
36	G16	C16	D9
37	C14	G20	A12
38	D13	C20	E8
39	B17	D13	D11
40	E16	E16	E18
41	E18	A14	A10
42	F19	C18	H17
43	E14	F5	D13
44	B13	D15	G8
45	H15	H7	C8
46	C18	E14	E14
47	D15	F19	F11
48	A14	B13	D7

49	B15	G16	H11
50	F15	E18	G10
51	G14	H13	F13
52	C6	D5	C20
53	E6	C6	F7
54	C8	B7	H15
55	F9	C8	F17
56	F13	F7	F15
57	F7	G10	H9
58	H13	G8	C6
59	F5	B17	G12
60	D7	E8	B19
61	C12	F9	E12
62	B9	H9	B1
63	G8	F13	D17
64	D9	D7	G14
65	E8	G6	B7
66	E12	B11	C12
67	G10	H11	B15
68	A8	C12	C10
69	E10	E6	G16
70	F11	A12	E16
71	H9	E10	B13
72	H11	G12	A8
73	B7	E12	C18
74	C10	D9	B9
75	D11	A8	B11
76	G6	B9	C14
77	B11	D11	E6
78	A12	C10	D15
79	G12	F11	C16
80	A10	A10	E10

Table S5. Chips rank-ordered by increasing average surprisal, based on data from the free-choice color-labeling task (See **Figure 3B**).

SI-Section 6: Munsell vs. CIELAB results

The color-naming data were obtained with chips defined by the standard Munsell array. As with all color-ordering systems, the Munsell system suffers some non-uniformities (7). To ensure that the results were not attributed to the peculiar defects of the Munsell system, we analyzed only those data for 24 color chips that sample the CIELAB color system evenly. The results show the same pattern: warm colors are associated with higher average surprisal compared to cool colors (**Figure S12**).

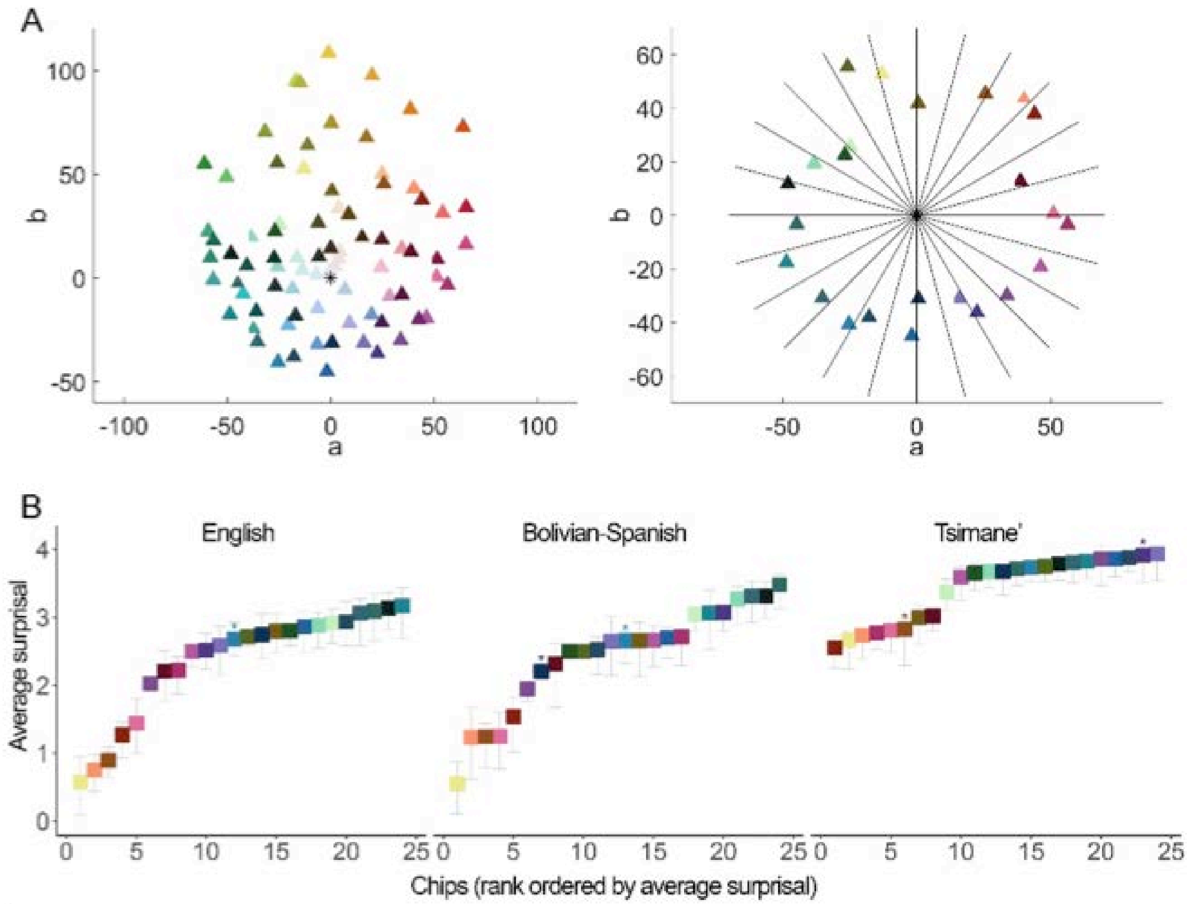

Figure S12. Color chips rank-ordered by their average surprisal (computed using equation 1), for Tsimane', Bolivian-Spanish and English, using only data for the 24 chips that uniformly sample the CIELAB color space. **A.** The 80 Munsell chips used in the color-naming experiment, plotted in the CIELAB space (left panel) and the subset of the chips that uniformly sample the CIELAB space (right panel). **Table S2** indicates the Munsell values for the 24 chips. The 24 chips were identified using an algorithm: first, the Munsell chips were projected into the CIELAB space, which was divided into 24 equal hue sectors; the chip within each sector that had chroma (saturation) value closest to 50 was selected. This procedure produced 24 chips that were roughly equal in saturation and that sampled the CIELAB space evenly around the hue circle. **B.** For all three languages, average surprisal was lower for warm colors compared to cool colors, for the subsampled chips. Spearman correlations: English-Spanish 0.74; Spanish-Tsimane' 0.43; English-Tsimane' 0.62.

SI-Section 7: Information-theoretic analysis & analysis with non-uniform prior

Equation (1) in the main text takes into account two factors: the probability P(w|c) that a given word will be produced to label the chip in question, and the log probability P(c|w) that a listener will correctly recover the chip in question from the word. As a result, both the consistency across the population in the words used for a given chip and the sampling density of the color space will impact estimates of average surprisal. For example, in English, a card painted with turquoise will have relatively high average surprisal (low communication efficiency) because there will be considerable variability in how the chip is labeled (green, blue, turquoise, cyan) and many other color chips could be labeled with these words. Conversely, a chip painted with focal red will have low surprisal (and high communication efficiency) because most people will use the term “red” to describe it, and few other chips will be labeled red.

The term P(c|w) is intended to represent the probability that a listener would choose a color chip c in response to color word w. We calculate P(c|w) from the color labeling data using Bayes rule:

$$P(c|w) = \frac{P(w|c)\,P(c)}{\sum_{c'} P(w|c')\,P(c')} \qquad \text{(equation SI-1)}$$

This calculation requires that we choose a prior P(c) over color chips. For the analysis above, we used a uniform prior over chips, in order not to bias the average surprisal scores toward favoring any colors in particular. This uniform prior was also used by Lindsey et al. (2015).

But if we believe that people are biased to talk about more salient colors, then using a uniform prior when calculating P(c|w) means that P(c|w) will not be a good approximation of the true probability that a speaker would choose a chip given a word. Here we show that using a salience-weighted prior does not affect the main result, that ranking color chips by average surprisal produces a universal warm-to-cool ordering.

We calculated the average surprisal of all color chips in the three datasets presented here and in the WCS data, this time using a prior P(c) proportional to the proportion of times that a color appears in a foreground object in the natural scene data. We argued above that the proportion of times a color appears in foreground objects is a measure of salience. The rank-ordered chips for all languages under this prior are shown in **Figure S13**. The overall informativity for English is 3.64 bits; for Spanish, 3.75 bits; for Tsimane', 4.76 bits. This analysis therefore qualitatively agrees with the one in the paper.

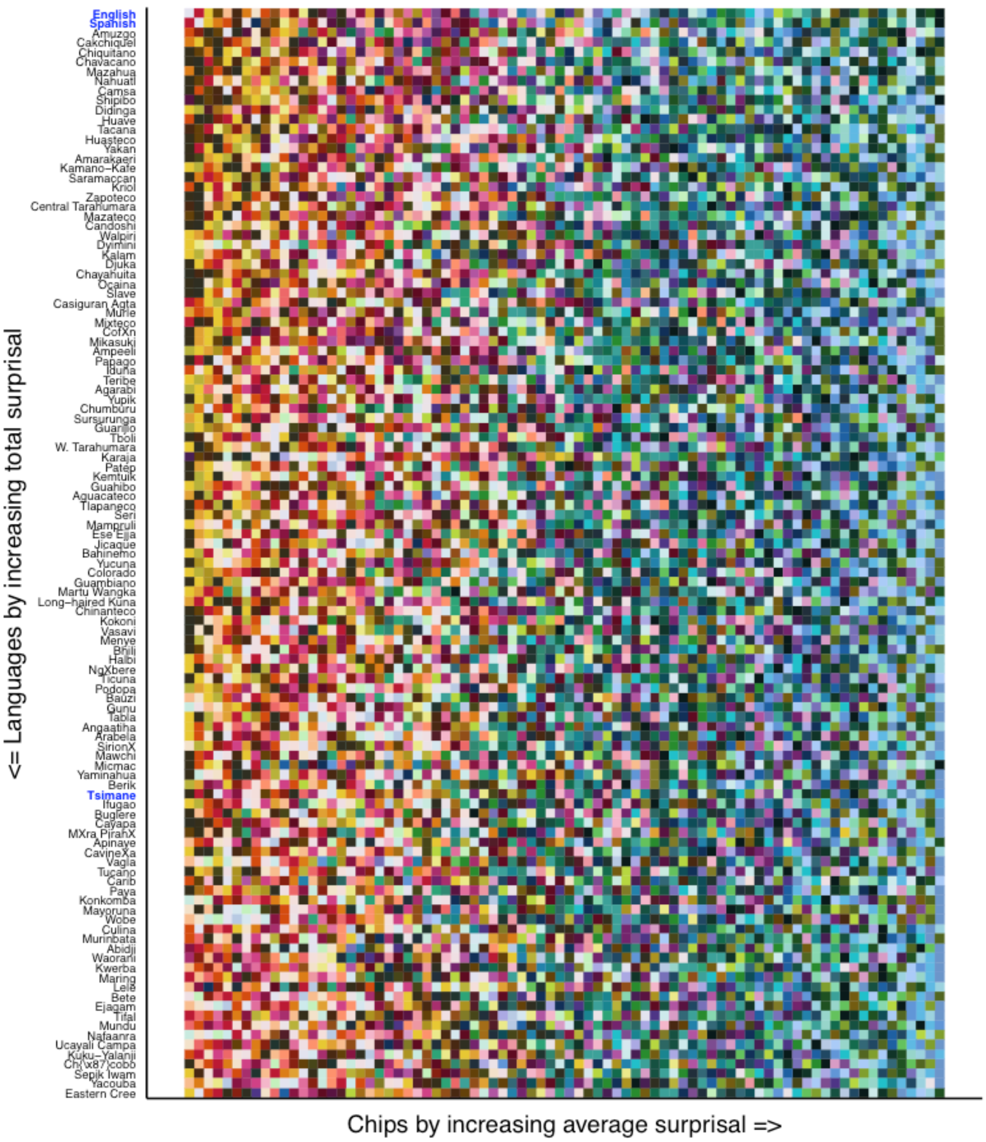

Figure S13. Color chips from the three datasets presented here and the WCS, rank-ordered by decreasing average surprisal under the non-uniform prior defined by the prevalence of colors in objects obtained in a large databank of natural images (compare with **Figure 4**).

SI-Section 8: Colors of objects identified in photographs

We analyzed the colors of “salient” objects identified in the Microsoft Research database of 20,000 natural images (8). This database, and similar databases obtained by collecting photographs posted on the internet, has been used to address a number of issues, including the assessment of artificial object recognition algorithms and the development of machine vision. The images in the Microsoft database were curated from over 200,000 photographs: human coders from Microsoft were tasked with identifying photographs depicting an object, and then within those photographs, the coders identified the objects using a bounding box. As part of our study, two people, ignorant of the purpose of our study, subsequently identified within the bounded areas of the photographs those pixels that comprised the object: regions of each image were traced using photoshop to create masks that contained the object and the background. The objects within the photographs were further subdivided into naturally colored and un-naturally colored categories. Using custom MATLAB scripts, the chromaticities of the pixels identified by the masked regions were then projected onto an equiluminant plane of the CIELUV color space within which we also projected the 80 Munsell color chips. The color of each pixel was then classified as one of the 80 Munsell colors used in the color-naming experiments (the Munsell color closest to the pixel color, defined using CIE xy chromaticity coordinates). For each of the 80 colors, we then determined the probability that the color would be found among the object pixels versus among the background pixels by computing: [(number of pixels of given color in objects – number of pixels of given color in backgrounds)/(number of pixels of given color in objects + number of pixels of given color in backgrounds)]. The correlations shown in **Figure 5** are maintained across the three languages (**Figure S14**).

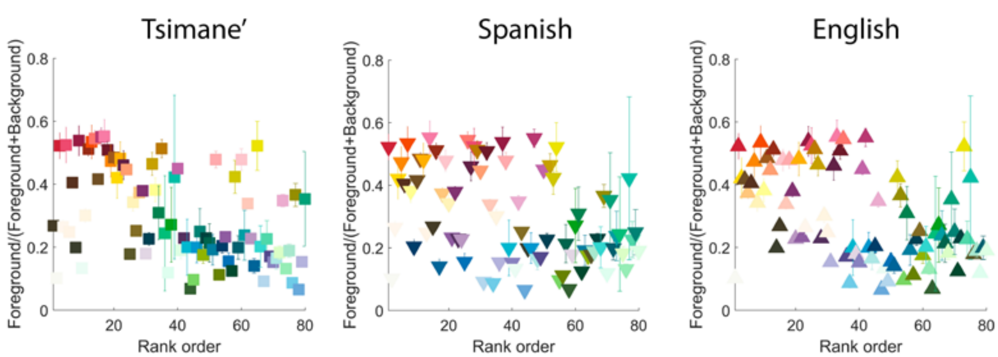

Figure S14. The color statistics of scenes containing objects predicts the average surprisal of colors. Objects in the Microsoft Research Asia (MRSA) database of 20,000 natural images were identified by human observers who were blind to the purpose of our study (see ref (29)). The colors of the pixels in the images were binned into the 80 colors defined by the Munsell chips used in the behavioral experiments (across the images there were 9.2x108 object pixels and 1.54x109 background pixels). The y-axis shows the probability of an “object”pixel having a given color, calculated as: [(number of pixels of given color in objects)/(number of pixels of given color in objects + number of pixels of given color in backgrounds)]. The three languages were not significantly different from each other (Tsimane’: slope = -0.003, Rho = -0.47; $p=1x10^{-5}$; Bolivian-Spanish: slope = -0.0025, Rho = -0.4; $p=3x10^{-4}$; English: slope = -0.003, Rho = -0.48; $p=6x10^{-6}$). Error bars show 95% C.I. computed through bootstrapping: the 20000 images were sampled with replacement to create 1000 sets, on which we performed the statistics.

We also compared the colors of objects with behavioral relevance to trichromatic primates with the communication efficiency of the colors (**Figure S15**). The data on the color statistics of the objects and backgrounds for this analysis were obtained using a spectroradiometer, thus they provide accurate representations of scene radiance, uncorrupted by the camera technology. These results confirm our main conclusions, showing that warm colors tend to have lower average surprisal than cool colors. Note that the spectral data analyzed in **Figure S15** (and analysis of physiological data from trichromatic non-human primates (9)) have been used to explain why trichromatic primates have relatively good discrimination of red versus green. But until now it has been assumed that categorization is equally good for warm versus cool. We show that this assumption is not valid: warm colors are subject to lower average surprisal compared to cool colors. This finding suggests a new explanation for the origin of the fundamental color category distinction between warm versus cool—that the distinction between warm and cool arose *because* of an asymmetry in the efficiency with which we communicate these colors. This explanation is not tautological, but rooted in the way the color-vision system is deployed for behavior.

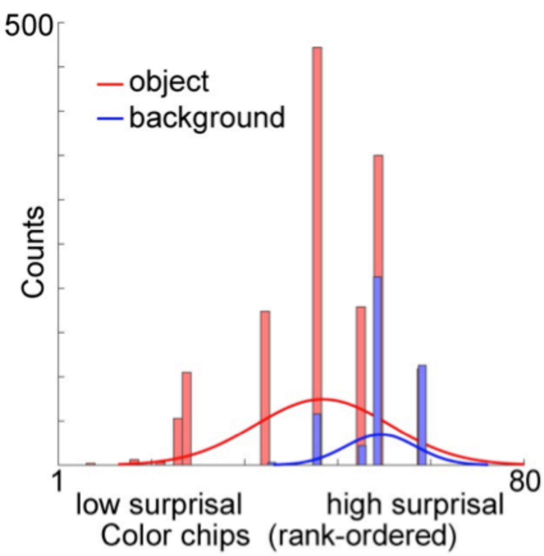

Figure S15. Colors associated with objects tend to have lower surprisal than colors associated with backgrounds, using calibrated spectral data (31). Spectral measurements from Regan et al (2001), obtained for objects that monkeys care about and objects that monkeys do not care about, were binned into the 80 Munsell chips. The histogram shows the surprisal for the distribution of samples identified as either "objects" or "backgrounds". The two distributions are significantly different (t-test, $p=10^{-58}$).

We are aware that prior work has attempted to draw correlations between color statistics in the natural environment and color categories (10). But this work has not incorporated any information about the behavioral relevance (to humans) of the colors. This is a crucial part of the present report. It is already well established that natural images have a bias for warm and cool colors (11-14), and the brain is adapted to these statistics (15-17). What we discovered is that warm colors have lower surprisal compared to cool colors, which is consistent with the new idea that it is the behavioral relevance of the colors, not simply their distribution in the natural world, that gives rise to the fundamental warm/cool color categories.

The images contained in the Microsoft database were undoubtedly taken using many different cameras under a range of different conditions and camera settings. We do not consider these images to be accurate representations of the color statistics of the objects depicted in the photographs; the images are simply useful for us to test the hypothesis about the color statistics of things that humans call objects (in this case, the objects are defined in the context of specific photographs) and the communicative efficiency of the colors associated with those objects. That the color statistics associated with any object depicted in a photograph deviates from the color statistics of the object viewed in the real world is not a concern here, because we are not asking about the faithfulness of the camera technology. Nonetheless, the analysis in **Figure S15** helps forge the link between our conclusions and the chromatic statistics of objects in the world.

One might ask why we bothered to conduct an analysis of the images in the Microsoft database given the availability of the spectral measurements from Regan et al (2001). The answer is that the Microsoft data base: (1) identifies objects using responses provided by human observers (not monkeys); (2) includes a much larger sample of objects, of a much wider array of object types; and (3) is a database used in machine vision/object-recognition algorithms (and is not unlike other photographic databases used for these purposes), so documenting the color statistics within this database is of independent value. Although we underscore that the spectral measurements of real objects estimated from the colors measured in the photographs are very likely inaccurate, because the cameras do not capture the full spectral content of the scene and often employ a number of compression and distortion algorithms implemented in order to render the photographs more appealing, it is noteworthy that color naming of objects seen in the real world and color naming of photographs of the same objects are highly correlated. Nonetheless, we need not invoke this correlation because we are simply interested in knowing whether there is any correlation between what a human observer calls "an object" and the color of it, regardless of what the object is (and whether it is in the real world or in a (poorly calibrated) photograph).

Prior work has addressed the relationship between the chromatic sensitivity of the photoreceptor pigments and natural scene statistics (18) or facial complexion (19). An analysis of photoreceptor responses may show how the visual system achieves sensitivity to the warm-cool chromatic axis, but it does not uncover the important asymmetry in communicative efficiency to warm versus cool colors, or the impact of culture, that we document here.

SI-Section 9: Use of color terms in a contrastive-labeling task

To assess the significance of between-language differences in likelihood of using a color word, we fit a mixed effect logistic regression predicting, for each trial, whether a color word was used. We included a fixed effect of language (English or Tsimane'), random intercepts for participant and object with a random slope by language for object. We found a significant effect of language (beta=-5.22, z=-5.88, $p<.0001$) such that Tsimane' speakers were less likely to use a color word, analyzing only trials in which the same head noun was used across the two similar items. The effect held even looking at only participants who used at least one color word or adjective (beta=-2.82, z=-4.54, $p<.0001$). This controls for the possibility that some participants may have understood the task as to label only the head noun, and not any distinguishing modifiers.

We performed a separate version of the experiment with a different group of 27 Tsimane' adults (mean age: 34.5 years; SD: 16.2 years; range 18-74; 22 females), in which the pairs of contrasting objects were presented at the same time. The contrasting color feature was even more apparent than when the objects were presented one at a time; the results of this experiment confirmed the conclusions drawn from the sequential task.

SI-Section 10: Tsimane' participants' knowledge of Spanish

As part of our testing procedure in Tsimane', we assessed participants' knowledge of Spanish words by asking them to translate 11 common Spanish objects into Tsimane' (e.g., perro ("dog"), rio ("river"), casa ("house")). The number of correct translations was coded numerically from 0 to 11, providing a rough estimate of their exposure to Spanish. To avoid inflated scores from participants who may have overheard the Spanish words while waiting for their turn, we used two different lists.

List 1: Perro (dog) hermano (brother) sal (salt) puerta (door) cabeza (head) vibora (snake) remo (oar) estómago (stomach) venado (deer) techo (ceiling) estrella (star)

List 2: río (river) diente (tooth) flecha (arrow) casa (house) negro (black) águila (eagle) choclo (corn) selva (jungle) pared (wall) pierna (leg) huevo (egg)

For the 58 participants that performed the free-choice task, the mean number of correct answers was 7.5 / 11, with only 3 getting all 11, across the seven villages where we tested. For the 41 participants that performed the fixed-choice task, the mean number of correct answers was 9.4 / 11, with 16 getting all 11 correct (9 of these were bilinguals from various villages, but tested in San Borja at CBIDSI; the other 32 were tested in three villages). For the free-choice task, all of the color words that we encountered were native (non-borrowed) Tsimane' except (a) the word for brown (cafedyeisi / chocoratedyeisi, borrowed from Spanish) and (b) azul the Spanish word for "blue", used by one participant.

Analysis of the relation between exposure to Spanish and color communication efficiency. To compare each participant's Spanish score with their efficiency of color-term usage we modified our measure of informativity of a color system to quantify the informativity of each individual speaker. To do so, we relied on equations (1) and (3) from Section 1. As before, the probability of selecting a chip given a word, $P(c|w)$, is computed using data from all participants. However, for the analysis in this section we compute a participant's probability of saying a word given a chip, $P(w|c)$, from the data only for that participant. That is, $P(w|c)$ for a participant is a conditional distribution with probability 1 on the word chosen by the participant given a chip, and 0 on all other words. This analysis quantifies how uncertain a random member of the population would be about color chips given the color words produced by an individual. If an individual uses color words consistently and similarly to the overall community, then the population's uncertainty about the intended color chips will be low, and we can say the individual's color language is highly informative. If an individual uses color words inconsistently and idiosyncratically, then the population's uncertainty about intended chips would be high, and her language would be less informative. **Figure S16** shows the relation between knowledge of Spanish and individual uncertainty computed this way. Using the data from the free-choice labeling task, we found a negative correlation between these two variables ($r=-0.318$; $t=-2.826$, df=71; p=0.006), suggesting that increased knowledge of Spanish results in a color word choice that reduces the population's uncertainty about the color chip being communicated.

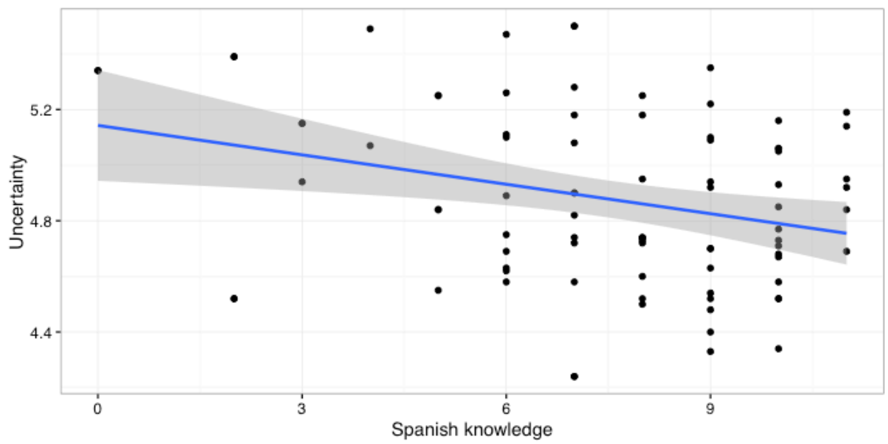

Figure S16. Relation between Spanish score (measure from 0 to 11), and the uncertainty in the population given each speaker's color word choices.

Although this analysis reveals a significant correlation between exposure to Spanish and communication efficiency, these effects could be driven by participants' age and/or education (which may both increase participant's knowledge of Spanish and their knowledge of color words). To test this possibility, we conducted a linear regression with conditional entropy as the dependent variable and age, education and knowledge of Spanish as the independent variables. Consistent with the first analysis, knowledge of Spanish was a significant predictor of conditional uncertainty. In contrast, age and education were not (**Table S6**).

	Estimate	Std. Error	t value	Pr(>|t|)	
Intercept	5.0175	0.143	35.173	<0.001	***
Education	-0.0002	0.014	-0.016	0.9875	
Age	0.0038	0.003	1.254	0.2134	
Spanish	-0.0343	0.013	-2.561	0.0123	*

Table S6. Knowledge of Spanish, but not age or education, predicted conditional entropy.

SI References

1. Berlin B & Kay P (1969) Basic color terms: their universality and evolution. *Berkeley, CA: University of California Press*.
2. Roberson D, Davidoff J, Davies IR, & Shapiro LR (2005) Color categories: evidence for the cultural relativity hypothesis. *Cogn Psychol* 50(4):378-411.
3. Lindsey DT, Brown AM, Brainard DH, & Apicella CL (2015) Hunter-Gatherer Color Naming Provides New Insight into the Evolution of Color Terms. *Curr Biol* 25(18):2441-2446.
4. Saunders BAC & vanBrakel J (1997) Are there nontrivial constraints on colour categorization? *Behav Brain Sci* 20(2):167-+.
5. Everett DL (2005) Cultural constraints on grammar and cognition in Pirahã. *Current Anthropology* 46(4):621-646.
6. Roberson D, Davies I, & Davidoff J (2000) Color categories are not universal: replications and new evidence from a stone-age culture. *J Exp Psychol Gen* 129(3):369-398.
7. Kuehni RG (2013) Color, An Introduction to Practice and Principles, 3rd Edition. *Hoboken, New Jersey: Wiley*:100.
8. Liu T, Sun J, Zheng N-N, Tang X, & Shum H-Y (2007) Learning to Detect A Salient Object. *Proc. IEEE Cont. on Computer Vision and pattern Recognition (CVPR), Minneapolis, Minnesota, 2007*.
9. Xiao Y, Kavanau C, Bertin L, & Kaplan E (2011) The biological basis of a universal constraint on color naming: cone contrasts and the two-way categorization of colors. *PLoS One* 6(9):e24994.
10. Yendrikhovskij SN (2001) Computing color categories from statistics of natural images. *Journal of Imaging Science and Technology* 45:409-441.
11. Vrhel MJ, Gerson R, & Iwan LS (1994) Measurement and Analysis of Object Reflectance Spectra. *Color Research and Application* 19(1):4-9.
12. Nascimento SM, Ferreira FP, & Foster DH (2002) Statistics of spatial cone-excitation ratios in natural scenes. *J Opt Soc Am A Opt Image Sci Vis* 19(8):1484-1490.
13. Webster MA, Mizokami Y, & Webster SM (2007) Seasonal variations in the color statistics of natural images. *Network* 18(3):213-233.
14. Webster MA & Mollon JD (1997) Adaptation and the color statistics of natural images. *Vision research* 37(23):3283-3298.
15. Lafer-Sousa R, Liu YO, Lafer-Sousa L, Wiest MC, & Conway BR (2012) Color tuning in alert macaque V1 assessed with fMRI and single-unit recording shows a bias toward daylight colors. *J Opt Soc Am A Opt Image Sci Vis* 29(5):657-670.
16. Conway BR (2014) Color signals through dorsal and ventral visual pathways. *Vis Neurosci* 31(2):197-209.
17. Lafer-Sousa R, Hermann KL, & Conway BR (2015) Striking individual differences in color perception uncovered by 'the dress' photograph. *Curr Biol* 25(13):R545-546.
18. Regan BC, *et al.* (2001) Fruits, foliage and the evolution of primate colour vision. *Philosophical Transactions of the Royal Society of London - Series B: Biological Sciences* 356(1407):229-283.
19. Changizi MA, Zhang Q, & Shimojo S (2006) Bare skin, blood and the evolution of primate colour vision. *Biol Lett* 2(2):217-221.