This is an original manuscript of an article published by Taylor & Francis in *Cognitive Neuropsychology* on April 26, 2019. Available online: doi.org/10.1080/02643294.2019.1604502.

Communicative need in color naming

Noga Zaslavsky,[a,b,✉] Charles Kemp,[c] Naftali Tishby,[a,d] and Terry Regier[b,e]

[a]Edmond and Lily Safra Center for Brain Sciences, The Hebrew University, Jerusalem 9190401, Israel; [b]Department of Linguistics, University of California, Berkeley, CA 94720 USA; [c]School of Psychological Sciences, The University of Melbourne, Parkville, Victoria 3010, Australia; [d]Benin School of Computer Science and Engineering, The Hebrew University, Jerusalem 9190401, Israel; [e]Cognitive Science Program, University of California, Berkeley, CA 94720 USA

ABSTRACT

Color naming across languages has traditionally been held to reflect the structure of color perception. At the same time, it has often, and increasingly, been suggested that color naming may be shaped by patterns of communicative need. However, much remains unknown about the factors involved in communicative need, how need interacts with perception, and how this interaction may shape color naming. Here, we engage these open questions by building on general information-theoretic principles. We present a systematic evaluation of several factors that may reflect need, and that have been proposed in the literature: capacity constraints, linguistic usage, and the visual environment. Our analysis suggests that communicative need in color naming is reflected more directly by capacity constraints and linguistic usage than it is by the statistics of the visual environment.

KEYWORDS

information theory; color naming; categorization; semantic typology

1. Introduction

Color naming varies widely across languages. At the same time, this variation is constrained, and certain universal tendencies of color naming recur across unrelated languages (e.g. Berlin and Kay, 1969; Lindsey and Brown, 2006). Figure 1 shows the color naming systems of four languages, illustrating this variation. As can be seen, both the number of terms and their extension vary across languages, but it is also the case that some cross-language commonalities can be found, such as the existence of terms roughly corresponding to English "red" and "yellow".

Why do the color naming systems of the world's languages vary as they do? Why do we see these systems and not other logically possible ones? Broadly speaking, three classes of explanation have been proposed, emphasizing color perception, communicative need, or both, as illustrated in Figure 2. Traditionally, cross-language variation has been explained largely in terms of perception (e.g. Kay and McDaniel, 1978, see Figure 2a). On this view, universal tendencies in color naming are relatively direct reflections of universals in color perception. Early work in this tradition did note in addition the apparent influence of cultural forces such as level of technological development, including dye technology, in determining the complexity of the color lexicon, but these ideas were not pursued in depth and were instead presented as "plausible speculation" (Berlin and Kay, 1969, pp. 16-17). The influence of

✉ noga.zaslavsky@mail.huji.ac.il

Cite as: Noga Zaslavsky, Charles Kemp, Naftali Tishby, and Terry Regier (2019). Communicative need in colour naming. *Cognitive Neuropsychology*. DOI: 10.1080/02643294.2019.1604502.

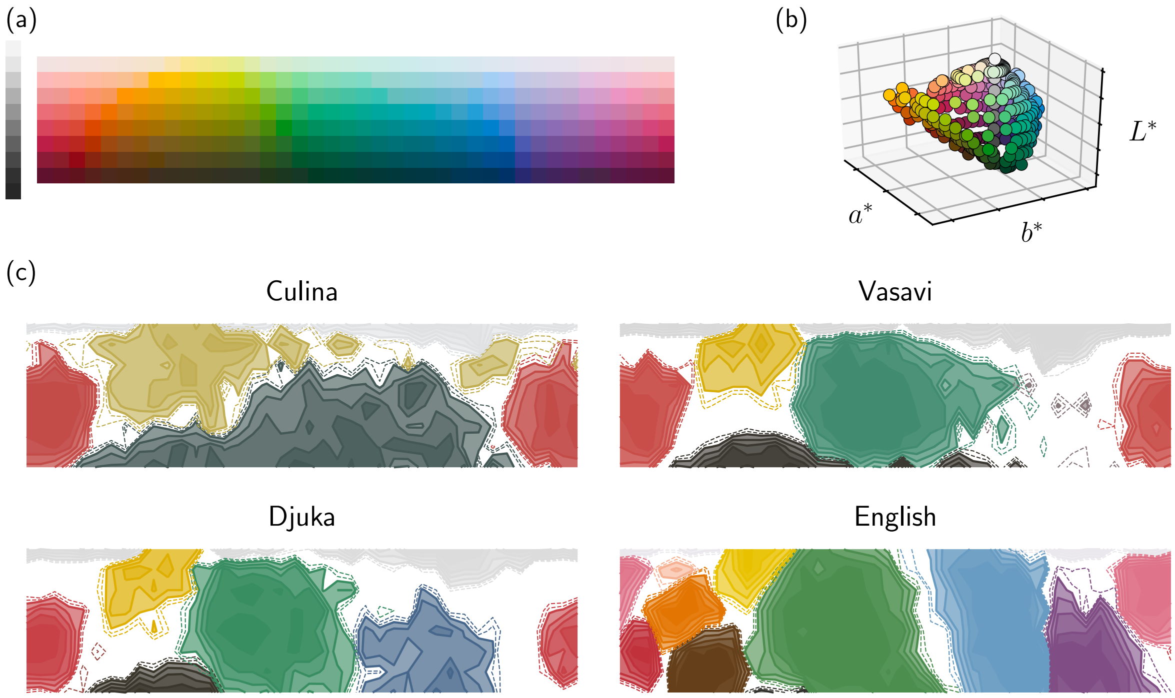

Figure 1. (a) A standard color naming stimulus grid, containing 320 color chips and 10 achromatic color chips (leftmost column). Columns correspond to equally spaced Munsell hues, rows correspond to equally spaced Munsell values (levels of lightness), and each chip is at the maximum available saturation (colorfulness) for that hue/lightness combination. (b) The 330 color chips re-plotted in the CIELAB perceptual color space, in which Euclidean distance between nearby colors is roughly correlated with perceptual dissimilarity. L^* corresponds to lightness, and hue and saturation are encoded in polar coordinates in the (a^*, b^*) plane. The chips are not evenly distributed in this space. For example, chips in the yellow region are exceptionally highly saturated (colorful) and therefore protrude farther outward away from other colors. This uneven distribution highlights presumably universal perceptual structure that may shape color naming across languages. (c) Examples of color naming systems from four languages in the color naming dataset we consider here, plotted against the stimulus grid. Each plot shows the contours of the naming probabilities for each term in the language. The naming probabilities for each color term are depicted in the color corresponding to the centroid for that term. Solid lines correspond to level sets of 50% and above, and dashed lines correspond to level sets of 40% and 45%.

communicative forces was later explored via multi-agent simulations (e.g. Steels and Belpaeme, 2005; Dowman, 2007; Loreto et al., 2012), and a recent elaboration of these ideas has suggested concrete roles for both perception and communicative need: specifically it has been proposed that color naming reflects perceptual structure as partitioned for communicative purposes (e.g. Jameson and D'Andrade, 1997; Komarova et al., 2007; Regier et al., 2007, see Figure 2b). In particular it has been proposed that color naming across languages may be shaped by the need for *efficient communication*: the need to communicate about color precisely, but at minimal cognitive cost (e.g. Lindsey et al., 2015; Regier et al., 2015). Recently Gibson et al. (2017) pursued this progression of thinking to its logical extreme, proposing that communicative needs do not merely modulate an effect of perceptual structure — but rather that communicative needs themselves govern the character of color categories (see Figure 2c), a hypothesis they situated as an "alternative" (p. 10785) to accounts based on perception. Here, we argue that need and perception should both be taken into account, in line with earlier efficiency analyses (back to Figure 2b). We review a principled theoretical framework for achieving this integration of perceptual structure and communicative need (Zaslavsky et al., 2018), and we use that framework to evaluate the character and role of communicative need in color naming.

The empirical basis for our evaluation is a set of color naming systems from 111 languages. These systems were drawn mainly from the World Color Survey (WCS: Kay et al., 2009), which contains color naming data from 110 languages of non-industrialized societies, with respect to the stimulus grid shown in Figure 1a. In addition, we consider color naming data from American English (Lindsey and Brown, 2014) which were collected with respect to the same stimulus grid. We refer to this joint dataset as the WCS+ dataset.

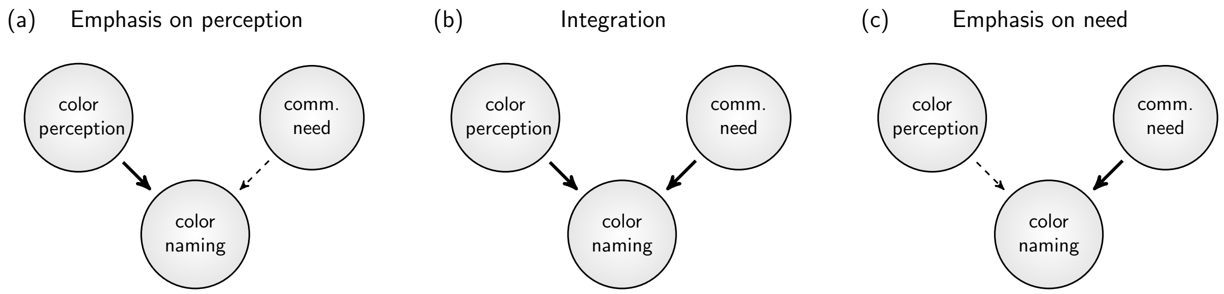

Figure 2. Color naming may be shaped by color perception, communicative need, or both.

The remainder of this paper proceeds as follows. In section 2, we review recent evidence suggesting that communicative need – together with perceptual structure – plays an important role in shaping color naming across languages. In section 3, we review a recent computational model that integrates communicative need and perceptual structure, and that accounts for color naming across languages in terms of an independent principle of efficiency. In this model, communicative need is formalized as a prior distribution over colors; however it is not yet clear how best to characterize this distribution. In section 4 we address this problem by presenting several estimation methods based on different factors that may reflect communicative need, specifically capacity constraints, linguistic usage, and the statistics of colors in the environment. Finally, in section 5, we evaluate these different factors by assessing how well the corresponding priors account for color naming data across languages, in the context of the model mentioned above.

2. The importance of communicative need

When considering the possible role of communicative need in shaping color naming, it is useful to distinguish two different kinds of need: *domain-level* need and *object-level* need (Kemp et al., 2018). Domain-level need is the communicative importance of a given domain, such as color, relative to other domains of human experience about which one may wish to communicate. For example, the observation that the introduction of dye technology may push a society or culture to develop a more fine-grained color lexicon, mentioned above, is an observation about domain-level need: with the advent of dyes, color as a domain presumably assumes greater cultural importance than it had previously, justifying greater complexity in this part of the lexicon. In section 3 we briefly discuss how domain-level need may be formalized in terms of the tradeoff between accuracy and complexity of the lexicon. Object-level need, in contrast, concerns how often one may need to communicate about particular objects within a domain — for example, within the domain of color, one may need to talk about certain colors more than others. It is this sort of object-level need that is naturally captured as a prior distribution over colors, and that is the primary focus of this paper.

As noted above, early accounts of color naming emphasized perception over (object-level) communicative need. Some justification for this stance is suggested by the fact that qualitative patterns of color naming across languages can be accounted for fairly well based only on perceptual structure, assuming a uniform prior over colors (e.g. Regier et al., 2015). However this leaves open the possibility that a better account of the data might be obtained with a non-uniform need distribution.

In line with this possibility, Gibson et al. (2017) argued that some colors are more useful than others for human purposes, and that the usefulness of particular colors is a major determinant of color naming across languages. Specifically, they argued for the greater usefulness of warm colors, relative to cool colors, and argued that this asymmetry in usefulness is reflected in patterns of color naming. They showed that across languages, color categories tend to support more precise communication for warm than for cool colors. They also examined color statistics in a large dataset of natural images and found that objects (as opposed to their visual backgrounds) tend to be warm-colored rather than cool-colored, in a parallel to the warm-cool asymmetry in language. They suggested on this basis that color naming across languages "reflects colors of universal usefulness" (p. 10785).

Zaslavsky et al. (2019) engaged this proposal, and argued for a somewhat different picture. They noted that the finding of a warm-cool asymmetry in language assumes a prior, and that the asymmetry vanishes under some well-motivated priors. They also found that the warm-cool naming asymmetry, when assessed using the same priors as Gibson et al. (2017), is present not only in natural color naming systems, but also in a set of artificial color naming systems that are based solely on perceptual structure, with no element of communicative need. These findings suggest that the warm-cool naming asymmetry, when it is found, cannot be taken as an unambiguous signature of communicative need. However this leaves open the possibility that there may be a different signature of need in the color naming data of the world's languages. Zaslavsky et al. (2019) proposed such a signature, based on the notion of a capacity-achieving prior (treated below in section 4.1), and found that natural color naming systems do indeed bear signs of communicative need beyond what would be predicted from perceptual structure alone. Thus, communicative need does appear to shape color naming in the world's languages.

A natural conclusion from the work just reviewed is that patterns of color naming may result from an integration of perceptual structure and communicative need. That conclusion leads to an important open question: what are the factors involved in communicative need, and how does need interact with perception in shaping color naming? The remainder of this paper addresses that question.

3. Integration of communicative need and color perception

The notion of efficient communication in color naming was recently formalized by Zaslavsky et al. (2018), building on earlier work by Regier et al. (2015), in a way that integrates perceptual structure and communicative need. Zaslavsky et al.'s proposal grounded the notion of efficient communication in an independently motivated information-theoretic principle, the Information Bottleneck (IB) principle (Tishby et al., 1999). On that basis, their proposal accounted to a large extent for the wide variation observed in color naming across languages, provided a theoretical explanation for the existence of soft color categories with graded membership, and synthesized previous accounts of color category evolution. For these reasons, we adopt the IB color naming model here as a framework within which different proposed sources of communicative need may be assessed.

The IB color naming model is based on a simple communication scenario between a speaker and listener, illustrated in Figure 3a. The speaker observes a color c drawn from a prior distribution $p(c)$ over colors in the environment $\mathcal{U}$, and wishes to communicate this color to the listener. The prior $p(c)$ reflects the communicative needs of the speaker, favoring certain colors over others (Kemp and Regier, 2012; Kemp et al., 2018). To account for perceptual uncertainty, it is assumed that the speaker does not have access to the exact color but rather to a noisy mental representation of it, m_c,[1] formulated as a Gaussian distribution centered at c over colors in the CIELAB perceptual space (Figure 1b). The speaker communicates this mental representation to the listener by producing a word w drawn from a shared lexicon $\mathcal{W}$, according to a naming distribution $q(w|c)$. The listener receives w and interprets this word by constructing a mental representation $\hat{m}_w$ that approximates the speaker's representation m_c.

According to the IB principle, the ideal speaker and listener are adapted to each other by jointly optimizing an information-theoretic tradeoff between the *complexity* of the lexicon and the *accuracy* of communication. This tradeoff is also illustrated in Figure 3a. Below, we lay out the IB formulations of complexity, accuracy, and their tradeoff.

In IB terms, a color naming distribution $q(w|c)$ is an encoder that compresses colors into words. As in rate-distortion theory (Shannon, 1959), the complexity of this encoder is measured by the information that the lexicon maintains about the speaker's representation, namely:

$$I(C;W) = \sum_{c\in\mathcal{U},w\in\mathcal{W}} p(c)q(w|c)\log\frac{q(w|c)}{q(w)}\,, \tag{1}$$

where $q(w) = \sum_c p(c)q(w|c)$. This informational complexity roughly corresponds to the number of bits

[1]For simplicity, since it is assumed that each color invokes a unique mental representation, we will treat c and m_c interchangeably when the distinction between them does not matter. For example, for any color naming distribution $p(w|c)$ or prior $p(c)$, it holds that $q(w|m_c) = p(w|c)$ and $p(m_c) = p(c)$.

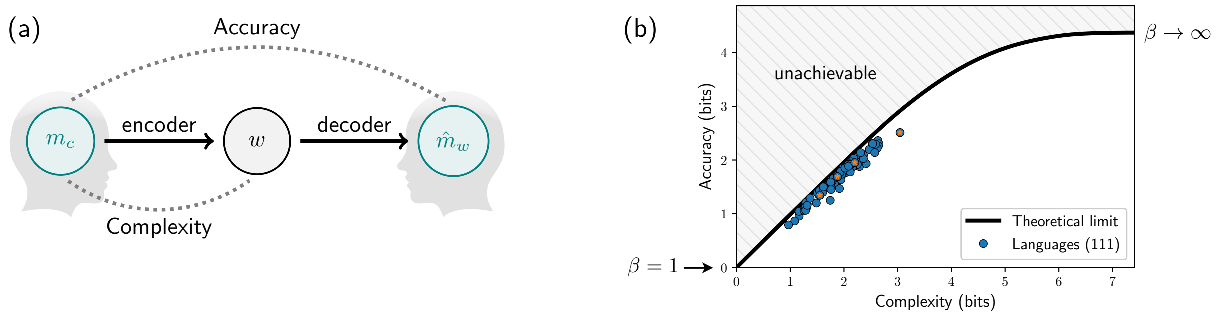

Figure 3. (a) The basic communication model. A color c is drawn from a prior distribution $p(c)$ that represents communicative need. The speaker observes c, mentally represents it by a distribution m_c, and communicates this representation to the listener by encoding it in a word w which is distributed according to an encoding naming distribution $q(w|c)$. The listener receives w and interprets (or decodes) it by constructing a mental representation $\hat{m}_w$. The complexity of the lexicon is determined by the encoder. The accuracy of the lexicon is determined by the similarity between the listener's and speaker's mental representations. (b) The theoretical limit of achievable complexity-accuracy tradeoffs, defined by the set of optimal IB systems, and the tradeoffs achieved by the color naming systems of the WCS+ languages. Accuracy is inversely related to the expected distortion (equation 2), such that maximal accuracy corresponds to zero distortion. All WCS+ languages achieve near-optimal tradeoffs. Orange stars correspond to the four languages shown in Figure 7, where they are ordered by complexity. Both figures are adapted from Zaslavsky et al. (2018).

that are required to represent the lexicon on average. Similar informational costs have also been proposed as measures for cognitive effort in other contexts (e.g. Ferrer i Cancho and Solé, 2003; Genewein et al., 2015; Tkačik and Bialek, 2016; Sims, 2016; Marzen and DeDeo, 2017).

The accuracy of communication is the extent to which the listener's interpreted representation is similar to the speaker's representation, or in other words the extent to which the distortion or discrepancy between these two representations is small. Since m_c and $\hat{m}_w$ are both distributions over color space, a natural distortion measure (Harremoës and Tishby, 2007) is the expected Kullback-Leibler (KL) divergence between them:

$$\mathbb{E}[D[m_c \| \hat{m}_w]] = \sum_{c \in \mathcal{U}, w \in \mathcal{W}} p(c) q(w|c) \sum_{u \in \mathcal{U}} m_c(u) \log \frac{m_c(u)}{\hat{m}_w(u)} \,. \tag{2}$$

There is necessarily a tradeoff between accuracy and complexity. Maximizing accuracy amounts to minimizing the distortion given by equation 2, which will be achieved when $D[m_c \| \hat{m}_w] = 0$, i.e. when $m_c \equiv \hat{m}_w$. This in turn will require a very complex lexicon, with a separate word for each color, so that each color can be communicated with perfect accuracy. On the other hand, minimizing complexity can be achieved by using a single word to describe all colors, but in this case accuracy will necessarily be low, i.e. communication will not be informative. The tradeoff between these two competing forces is given by the following equation, which is equivalent to the standard IB objective function:

$$\min_{q(w|c), \hat{m}_w} \; I(C;W) + \beta \, \mathbb{E}[D[m_c \| \hat{m}_w]] \,, \tag{3}$$

where the tradeoff parameter $\beta \geq 0$ controls how complexity and accuracy are balanced.

The optimal IB color naming systems, i.e. the systems that optimize equation 3 for different values of β, define the theoretical limit of achievable tradeoffs. Zaslavsky et al. (2018) evaluated this theoretical limit and found that the color naming systems in the WCS+ data are near-optimal in that they lie near this theoretical limit (Figure 3b). This suggests that languages may have evolved under pressure for information-theoretic efficiency. It can be seen that variation in the tradeoff parameter β accounts for much of the cross-language variation in the WCS+ data — meaning that different languages navigate the tradeoff between accuracy and complexity in different ways, while remaining near the theoretical limit of efficiency. It is natural to interpret β as capturing domain-level need, or the cultural importance of color as a domain in a given society (recall section 2): the more important it is to communicate accurately about color, the more it is justified to allow greater complexity to achieve that accuracy — and this tradeoff is exactly what β controls. This notion is captured memorably in the title of a paper that

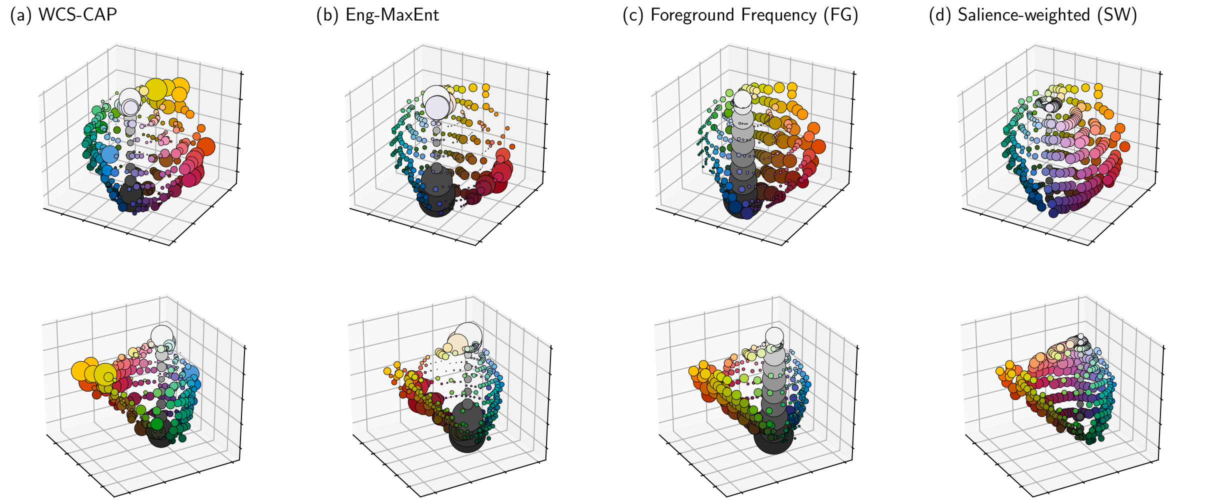

Figure 4. Illustration of how communicative need may interact with perceptual structure in shaping color naming. Each plot shows the 330 color chips from Figure 1b as circles in CIELAB space, where the size of each circle is proportional to the color's probability mass under four different priors defined in section 4. Each prior is a different distribution over perceptual space, which may give rise to different color naming systems.

described color naming in a society for which color is relatively unimportant: "We don't talk much about colour here" (Kuschel and Monberg, 1974); as would be expected, the color system of this language was found to be very simple, having only three basic color terms.

Importantly, the findings just reported were obtained with a specific non-uniform prior which is based on the notion of capacity-achieving priors (WCS-CAP, see section 4.1 for detail). It is not yet clear whether other well-motivated priors could provide a better account of the data.

In what follows, we systematically investigate the effect of different priors while keeping all the other components of the IB color naming model fixed. Preparatory to doing so, it may be useful to note how both perceptual structure and the prior influence the IB objective function. The irregular distribution of colors in perceptual space (Figure 1b) influences the accuracy term (equation 2), through m_c and $\hat{m}_w$. The prior $p(c)$ influences both terms of the IB objective function: complexity (equation 1) and accuracy (equation 2). Colors with higher communicative need, i.e. higher $p(c)$, will therefore be more dominant in the IB objective function (equation 3), and thus there will be greater pressure to communicate those colors efficiently. Figure 4 illustrates this concretely, by showing how different priors we explore in the next sections emphasize different parts of perceptual color space.

4. Characterizing communicative need

We explore three general classes of prior distribution, each derived from a different principle for inferring communicative need. First is the class of least informative priors. This class aims to infer a prior without making any assumptions about external forces that may shape communicative need. The second class is based on the idea that communicative need is reflected in linguistic usage. The third class is based on the assumption that communicative need is reflected in color statistics as encountered in the visual world, estimated from natural images.

4.1. Least informative priors

A natural approach to obtaining a prior distribution without any assumptions is by invoking the maximum entropy (MaxEnt) principle (Jaynes, 1982). The MaxEnt principle states that the most justified distribution is the one that maximizes uncertainty, measured in terms of entropy. In our setting, in its simplest form, this principle yields a uniform distribution over color chips. A uniform prior has been used before to account for color naming (e.g. Regier et al., 2015; Gibson et al., 2017), and thus we

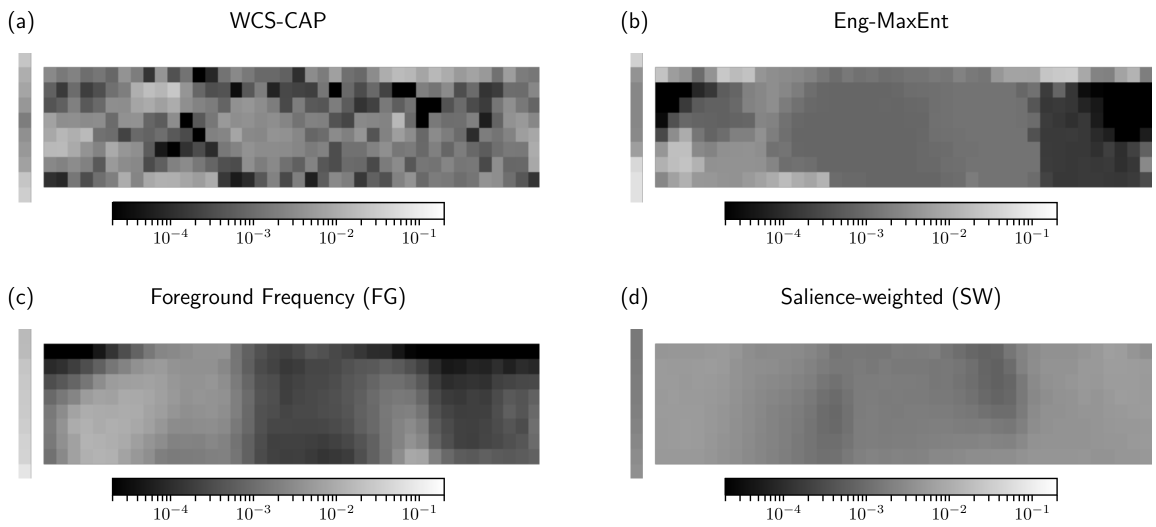

Figure 5. Prior distributions over the WCS grid. Each chip is colored according to its probability mass (log-scale).

consider it here as a baseline. However, it is not clear whether in fact all colors are equally needed for communication in natural settings.

An alternative approach (Zaslavsky et al., 2018) aims to infer the prior directly from naming data, without making specific assumptions about the forces that may shape communicative need. This approach is based on the capacity-achieving principle (Shannon, 1948). In information theory, a channel is defined by a conditional distribution (Cover and Thomas, 2006). Thus, any color naming distribution, $p(w|c)$, can be interpreted as a channel[2] that takes a color c as its input and outputs a word w. The maximal amount of information that can be transmitted over a channel is the channel's capacity, and the ideal prior for that channel is called a capacity-achieving prior (CAP). In our setting, the CAP for a given naming distribution maximizes the amount of information the lexicon conveys about the observed color. Formally, it is defined by

$$p_{\text{CAP}}(c) = \underset{p(c)}{\operatorname{argmax}}\, I(C;W)\,, \tag{4}$$

and can be obtained using the Blahut-Arimoto algorithm (Blahut, 1972; Arimoto, 1972). Note that this CAP identifies a prior that maximizes complexity (equation 1) for a given naming system, in contrast to the IB principle in which complexity is minimized over all possible naming systems for a given prior. Although these principles are related, they are also importantly different: the capacity-achieving principle is an optimality criterion for the prior whereas the IB principle is an optimality criterion for the naming system.

Given a color naming distribution $p_l(w|c)$ for a specific language l, we can now obtain a CAP for that language, $p^{(l)}_{\text{CAP}}(c)$, which captures the pattern of communicative need for that language, inferred on the basis of the capacity-achieving principle. We then follow Zaslavsky et al. (2018) and average together the CAPs across languages l in the WCS+ dataset, to obtain a single universal prior.[3] The resulting prior, which we refer to as WCS-CAP, is shown in Figure 4a and Figure 5a. Because this prior is estimated from the WCS+ data, Zaslavsky et al. (2018) performed 5-fold cross-validation and showed that WCS-CAP does not overfit the data (see Table 1).

[2]This *naming channel* is internal to the speaker, and it is distinct from the *communication channel* between the listener and speaker. The latter takes as input the word produced by the speaker and outputs the word perceived by the listener. The communication channel is left implicit in Figure 3a because this channel is assumed to be noiseless — i.e., the listener observes the speaker's word unaltered.

[3]For compatibility with the analysis performed by Zaslavsky et al. (2018), we followed their regularization process and excluded fifteen languages from our quantitative evaluation (Table 1). We also repeated the evaluation process with all languages and obtained similar results; thus the regularization process does not influence our conclusions.

4.2. *Linguistic usage*

It seems likely that the frequency of use of particular words in natural communication may reflect important aspects of communicative need, and priors estimated from corpus frequencies have been used to account for cross-linguistic variation in semantic domains other than color (Kemp and Regier, 2012; Xu and Regier, 2014). However, a challenge for this approach is that it is not always clear how to infer a distribution over objects in the domain — colors, in our case — from corpus statistics, because corpus statistics provide frequencies only for words, and there are generally more objects in the domain (here, color chips) than there are words (color terms). Here we propose a general solution for this problem by applying the maximum entropy (MaxEnt) principle under constraints derived from corpus data.

Suppose we are given the naming distribution $p_l(w|c)$ for some language l, and we are also given word frequencies, $p_l(w)$, from a corpus for that language. For simplicity, assume that these word frequencies correspond only to cases in which these words are used for describing objects in the domain universe $\mathcal{U}$. Under this simplifying assumption, for $p_l(w|c)$ and $p_l(w)$ to be consistent with each other, it must hold that $\sum_c p_l(w|c)p(c) = p_l(w)$. This consistency requirement imposes a set of linear constraints on the prior, and of all the prior distributions that satisfy these constraints, we wish to select the one with maximal entropy, where entropy is defined by $H(C) = -\sum_c p(c)\log p(c)$. Formally, this gives the following optimization problem:

$$\begin{aligned} \max_{p(c)} \quad & H(C) \\ \text{subject to} \quad & \sum_{c\in\mathcal{U}} p_l(w|c)p(c) = p_l(w), \quad \forall w \in \mathcal{W}\,. \end{aligned} \tag{5}$$

This is a concave optimization problem, and can be solved using standard tools.[4]

In principle, this corpus-based MaxEnt approach can be applied on a language-specific basis, for every language for which $p_l(w)$ can be obtained. However, it is difficult to obtain such word frequencies for the WCS languages, because large representative corpora for these languages of non-industrialized societies do not exist. For English, in contrast, this approach is tractable because both naming data and corpus data exist. The English color naming data collected by Lindsey and Brown (2014) contain over 100 words used across participants in their free-naming experiment. However, most of these words were used by only a few participants (see Lindsey and Brown, 2014). These words tend to be either rare, in which case their corpus frequencies may not be reliable, or words that are used metaphorically, in which case their frequencies are more likely to reflect usages other than describing colors. To mitigate this problem, we based this prior on only the 11 basic color terms in English, which were used by all participants. We also obtained corpus frequencies for these 11 terms, as shown in Figure 6. The naming

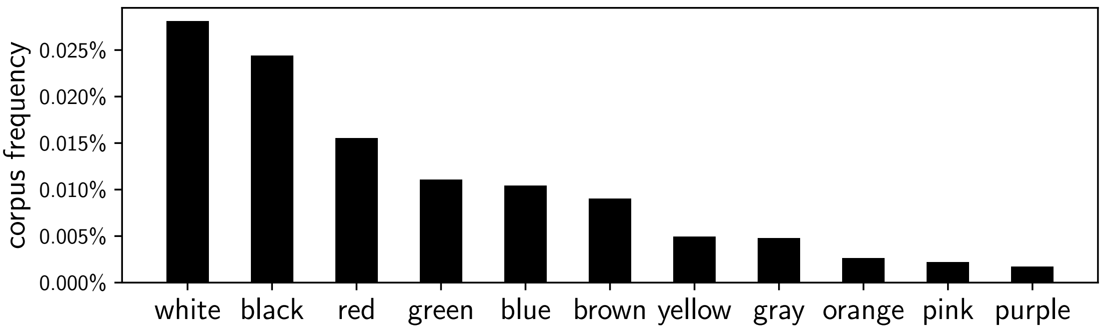

Figure 6. Frequencies of the 11 basic color terms in English (case insensitive) from the Google n-gram (Michel et al., 2011) American English dataset for the year 2008 with a smoothing factor of 3 (average across the three preceding years). Since the English naming data from Lindsey and Brown (2014) were collected in the USA, this is a reasonably compatible corpus.

[4]We used the python package `cvxopt` to solve this optimization problem. In general, it is possible that the feasible set would be empty, i.e. that there would be no prior that satisfies the constraints. However, this is not the case in our setting.

data and corpus frequency of each basic color term define the constraints in equation 5.

The resulting prior, Eng-MaxEnt, is shown in Figure 4b and Figure 5b. In contrast to WCS-CAP, this prior is estimated only from the English color naming data and English corpus statistics, and thus it is independent of the WCS languages. We explore Eng-MaxEnt as a proposed approximation to a universal prior, on the assumption that corpus statistics in English may be shaped in part by universal communicative forces. We leave the interesting question of language-specific differences in usage and communicative need for future research (but see Regier et al., 2016, for treatment of this idea in another domain).

4.3. Visual environment

A natural possibility is that communicative need may be shaped largely by the statistics of colors in the world (e.g. Yendrikhovskij, 2001; Gibson et al., 2017). If this is the case, then a prior derived from the distribution of colors in the environment should provide a good account of color naming. One way to approximate this distribution is from the statistics of colors in a large dataset of natural images. For example, Yendrikhovskij (2001) considered the total frequency of colors in a set of natural images. Gibson et al. (2017) also examined color frequencies in natural images, but they noted that not all occurrences may be equally relevant for estimating need. Instead, they took as their measure of communicative need what they called the "salience" of particular colors: specifically, the frequency of a color's appearance in objects that people tend to talk about, divided by the overall total frequency of that color. Here we consider these two approaches, and another that is based only on a color's frequency of appearance in foreground objects. This latter approach is based on the observation that if colors that appear in useful objects have greater communicative need, then this may hold regardless of the visual background of these objects.

To evaluate these different image-based approaches, we estimated (i) a prior based on the total frequency (TF) of colors; (ii) a prior based on the frequency of colors in foreground objects (FG); and (iii) a salience-weighted (SW) prior, similar to Gibson et al.'s approach but here based on the colors corresponding to all WCS chips whereas their analysis was based on a subset of these chips. Color frequencies were estimated from Microsoft's COCO dataset (Lin et al., 2014), which contains over 80,000 annotated images.[5] The images were processed as follows. First, to filter out black-and-white images, only images with colorfulness index (Yendrikhovskij et al., 1998) above 0.2 were considered. Approximately 3% of the images were excluded on this basis. Next, to avoid a bias toward large images, 50,000 pixels were randomly sampled from each image. These pixels were then converted to CIELAB coordinates (Figure 1b) and were classified as one of the WCS chips, or excluded if they were not close to any of the WCS chips.[6] Pixels with chroma less than the average chroma of pixels in the image were compared to the achromatic chips. Pixels with chroma above average were compared to the chromatic chips with closest lightness and hue values.

The resulting SW and FG priors are shown in Figure 4 and Figure 5. The TF prior is not shown because it is fairly similar to the FG prior.[7]

5. Results

We assess the three classes of priors discussed above by entering each prior into the IB color naming model, and evaluating how well the model with this prior accounts for the WCS+ data. We follow the same quantitative evaluation method used by Zaslavsky et al. (2018), which is based on two

[5] We considered the 2014 training dataset which contains 82,783 images. These images are annotated with object boundaries for objects from 80 different categories.

[6] Conversion from RGB to CIELAB coordinates was done with the `colorspacious` python package, using illuminant C. For the achromatic chips, only pixels with $\Delta E^2 = (L^*)^2 + (a^*)^2 + (b^*)^2 < 70$ were considered. For the chromatic chips, the comparison was based only on lightness and hue values, and pixels for which the square distance to the closest chromatic chip was greater than 400 were excluded. These thresholds were validated by manual inspection, to ensure that the converted pixels are indeed perceptually similar to the original ones.

[7] The TF and the FG priors have similar structure and both give the highest probability mass to the achromatic colors. However, the FG prior gives less weight to the achromatic chips than the TF prior does. In addition, according to the FG prior, warm colors have higher probability than cool colors, similar to the SW prior we estimated, and consistent with the salience data of Gibson et al. (2017).

goodness-of-fit scores:[8] (i) an inefficiency score, which measures the deviation from optimality of a given language's color naming system; and (ii) a dissimilarity score, which measures the dissimilarity in extension between a given language's color naming system and the corresponding optimal naming system predicted by the model. Lower values of these scores indicate a better fit to the data.

Table 1 shows the quantitative results based on these scores. WCS-CAP and Eng-MaxEnt achieve comparable scores, and outperform the other priors. A qualitative inspection of the results (Figure 7) shows that these priors predict slightly different solutions, but also agree to a large extent on the structure of the categories and resemble the actual systems. It is striking that Eng-MaxEnt — a prior that is derived only from English — is able to account so well for the WCS languages, which are from non-industrialized societies and the majority of which have fewer color categories than English. This result suggests that there are general patterns of communicative need that are shared across cultures, and that these patterns can be inferred directly from linguistic data. While it is possible that Eng-MaxEnt and WCS-CAP also reflect perceptual structure, the influence of perception on these priors would be indirect, mediated via language use (Winter et al., 2018). For completeness, we compared these results with those obtained by using a capacity-achieving prior estimated only from English naming data, and not those of any other language. This prior does not produce as good a fit to the actual data as do Eng-MaxEnt and WCS-CAP.

The relatively poor performance of the image-based priors is somewhat surprising, especially given that prior work (e.g. Yendrikhovskij, 2001; Griffin, 2006; Gibson et al., 2017) suggested that image statistics may play a central role in accounting for color naming. Looking more closely at the results from the image-based priors may help to explain this seemingly inconsistent outcome.

Consider first the TF and FG image-based priors. They achieve similar scores and both perform better than the SW prior and the uniform prior, but not as well as the priors based on linguistic data (WCS-CAP and Eng-MaxEnt). These results seem inconsistent with the findings of Yendrikhovskij (2001), who found that colors sampled from 630 natural images form clusters in color space that correspond roughly to known universal tendencies in color naming. However, Steels and Belpaeme (2005) found that categories generated by Yendrikhovskij's method are correlated with human color categories only slightly better than are categories derived from uniform sampling of colors.[9] In an attempt to more completely explore the apparent tension between our findings and those of Yendrikhovskij, we tried to replicate the findings of Yendrikhovskij (2001) using 1000 random images from the COCO dataset. Our analysis failed to replicate the qualitative results he obtained. This negative outcome could be due to the fact that we used a different set of images, or that the distribution of images in the COCO dataset is biased toward western cultures. However, there is also a further potential explanation for why the TF and FG frequencies do not perform well: they may not give good estimates of communicative need. Specifically, since most colors in natural images have low saturation (e.g. Hendley and Hecht, 1949; Steels and Belpaeme, 2005), the TF and FG frequencies are biased toward the achromatic chips. In our analyses, we excluded

Table 1. Evaluation of possible communicative need distributions.

Motivation	Data type	Prior	Inefficiency	Dissimilarity
Baseline	None	Uniform	0.24 ($\pm$0.09)	0.39 ($\pm$0.12)
Least informative	WCS+	WCS-CAP	**0.18** ($\pm$0.07)	0.18 ($\pm$0.10)
Linguistic usage	English naming & corpus data	Eng-MaxEnt	0.19 ($\pm$0.09)	**0.17** ($\pm$0.08)
Visual environment	Foreground freq. Total freq. Color salience	FG TF SW	0.21 ($\pm$0.08) 0.21 ($\pm$0.08) 0.25 ($\pm$0.09)	0.31 ($\pm$0.12) 0.34 ($\pm$0.14) 0.40 ($\pm$0.12)

Inefficiency and dissimilarity scores are as defined by Zaslavsky et al. (2018). Reported scores correspond to averages across languages ± 1 SD. Lower values are better, and the best scores are in boldface. Results for the two uninformative priors are from Zaslavsky et al. (2018), where the scores for WCS-CAP are averages over left-out languages in 5-fold cross-validation.

[8] These two measures correspond to ε_l and gNID respectively. See (Zaslavsky et al., 2018) for more detail.

[9] We thank Delwin Lindsey for drawing our attention to this connection.

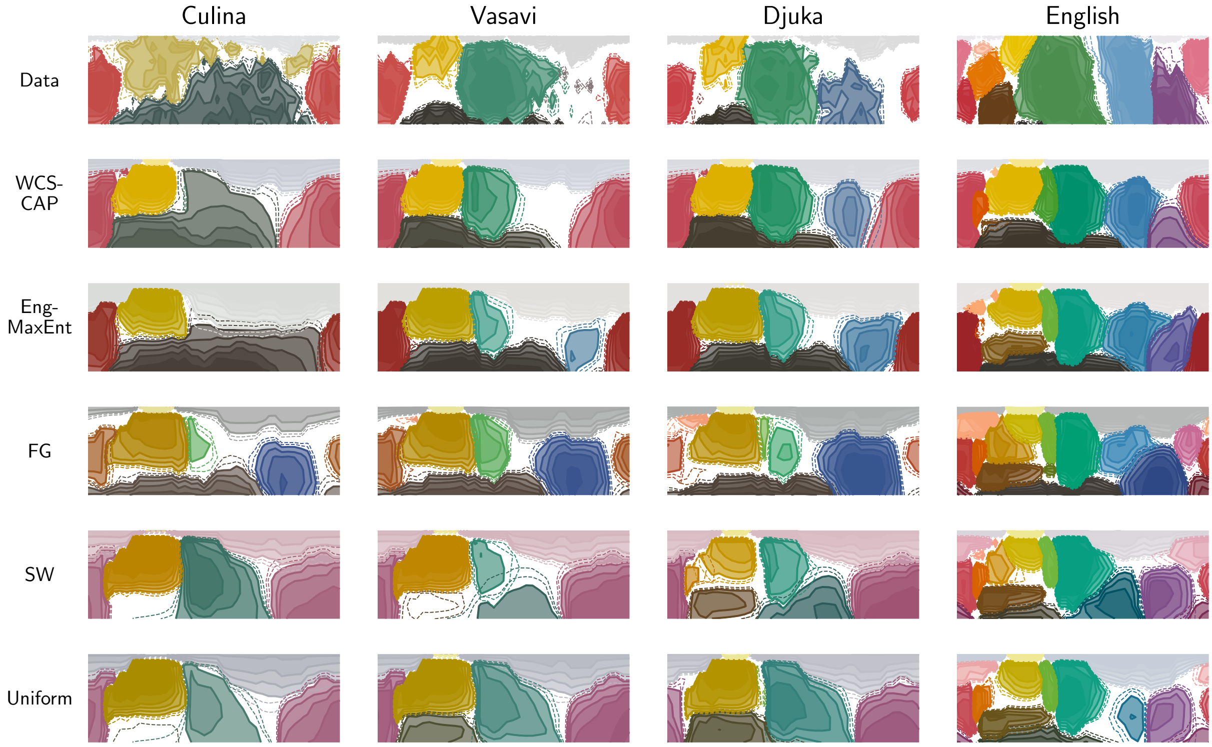

Figure 7. Contour plots of color naming systems from four languages (data row, same as Figure 1c) and the corresponding optimal systems which were predicted by the IB model under different priors. The variation shown for each model's prediction is caused by changes in the tradeoff parameter β that controls the location along the theoretical limit (see Figure 3b and section 3). Results for WCS-CAP and the uniform prior are from Zaslavsky et al. (2018).

colors that were not sufficiently close to any of the WCS chips, but the bias toward the achromatics seems inherent to the statistics of colors in images in general, prior to any exclusion or filtering: the density near the achromatic chips is much higher than the density near the chromatic chips. This implies that the TF and FG priors predict greater communicative need for desaturated colors. Such a tendency seems unlikely given that consensus in color naming, at least among English speakers, is positively correlated with chroma, such that highly saturated colors are named with highest consensus (Jraissati and Douven, 2018).

Consider now the SW prior. This prior is not biased toward desaturated colors. At the same time, it is closer to uniform than the other priors (Figure 5), it achieves scores similar to those of the uniform prior (Table 1), and it also predicts systems qualitatively similar to those predicted by the uniform prior (Figure 7). This suggests that the SW prior may be too close to uniform to accurately reflect communicative need.

6. Discussion

The possibility that both perceptual structure and communicative need may shape color naming has long been discussed in the literature. However perception has traditionally been the focus of much more attention, and was incorporated first in computational accounts of color naming, while communicative need remained an informal concept. Recently, this picture has started to change: the notion of communicative need has been cast formally as a prior over colors, and there is increasing evidence for the importance of this component. However, the factors that may characterize and shape communicative need have previously been only preliminarily explored. We approached this problem by exploring three major factors that may shape communicative need: capacity constraints, linguistic usage, and the visual environment. These factors were assessed within an independently motivated computational framework that integrates need and perception, and that predicts optimally efficient color naming systems on that basis.

Our findings may be summarized in two main points. First, we found that different patterns of communicative need, instantiated as different priors, give rise to quite different efficient color naming systems, given the same underlying perceptual structure. This finding further supports the idea that communicative need may have a substantial impact on color naming, beyond the influence of perception. Second, we found that of the priors we considered, those based on capacity constraints and linguistic usage provided the best fit to actual color naming systems observed across languages. These best-performing priors were estimated from linguistic data, whereas other priors — uniform and image-based priors — did not account for the data as well. This suggests that communicative need may be well-estimated by the statistics of linguistic usage (Kemp and Regier, 2012; Xu and Regier, 2014; Regier et al., 2016), rather than by the statistics of the visual world to which language refers.

The corpus-based maximum entropy method for estimating need that we have presented here is novel, to our knowledge, and seems noteworthy for two reasons. First, it addresses the challenge of inferring communicative need from corpus statistics with minimal additional assumptions, and it can therefore in principle be applied widely across semantic domains. Second, while its performance is comparable to that of the capacity-achieving prior based on multiple languages in our dataset, it achieves this based on data from a single language. This suggests that there are important aspects of communicative need that are shared across languages, and that this method can be used to infer them. At the same time, we are not committed to the notion of an entirely universal prior. An important direction for future research is to test how well this corpus-based maximum entropy approach generalizes across languages and across domains, and to determine how and why communicative need varies across cultures, environments, and languages, beyond the simplifying assumption of a universal prior that we have made here.

Our findings do not imply that communicative need is uninfluenced by the statistics of the visual environment. Instead, they suggest that any influence of visual environment may be distal, and that language use may be a more direct reflection of need. This is broadly consistent with Boas' (1911, p. 26) view that cross-language variation in semantic categories "must to a certain extent depend upon the chief interests of a people": on this view, while the environment may shape a people's interests, it is those interests that directly shape the semantic categories of a given language – and those interests are presumably expressed through patterns of language use. This suggests two linked processes of adaptation. In the case of color, color naming may have adapted to communicative need and the structure of perceptual color space — while need and perception may themselves have adapted to natural scene statistics (Shepard, 1994; Webster and Mollon, 1997), which may vary over time (Webster et al., 2007) and space (McDermott and Webster, 2012). Although we have focused here on forces that shape color naming, either directly or indirectly, it is also known that color naming may in turn shape color cognition and perception (Kay and Kempton, 1984; Roberson et al., 2000; Gilbert et al., 2006; Winawer et al., 2007; Bae et al., 2015). Given the many moving parts in this overall picture, we find it striking that a universal perceptual color space, and a universal prior based only on English usage, account for cross-language data as well as they do. Future research can usefully explore why this is the case, how far the universality extends, and when and under what circumstances language- and culture-specific forces dominate instead.

Acknowledgments

We thank Joshua Abbott for helpful discussions, and Delwin Lindsey and Angela Brown for kindly sharing their English color-naming data with us. This study was partially supported by the Gatsby Charitable Foundation (N.Z. and N.T.), and by the Defense Threat Reduction Agency (N.Z. and T.R.); the content of the study does not necessarily reflect the position or policy of the U.S. government, and no official endorsement should be inferred.

References

Arimoto, S. (1972). An algorithm for computing the capacity of arbitrary discrete memoryless channels. *IEEE Transactions on Information Theory*, 18(1):14–20.

Bae, G.-Y., Olkkonen, M., Allred, S. R., and Flombaum, J. I. (2015). Why some colors appear more memorable than others: A model combining categories and particulars in color working memory. *Journal of Experimental Psychology: General*, 144:744–763.

Berlin, B. and Kay, P. (1969). *Basic Color Terms: Their Universality and Evolution*. University of California Press, Berkeley and Los Angeles.

Blahut, R. E. (1972). Computation of channel capacity and rate-distortion functions. *IEEE Transactions on Information Theory*, 18(4):460–473.

Boas, F. (1911). Introduction. In *Handbook of American Indian Languages, Vol.1*, pages 1–83. Government Print Office (Smithsonian Institution, Bureau of American Ethnology, Bulletin 40).

Cover, T. and Thomas, J. (2006). *Elements of Information Theory (2nd Edition)*. Wiley-Interscience, Hoboken, NJ.

Dowman, M. (2007). Explaining color term typology with an evolutionary model. *Cognitive Science*, 31(1):99–132.

Ferrer i Cancho, R. and Solé, R. V. (2003). Least effort and the origins of scaling in human language. *Proceedings of the National Academy of Sciences*, 100(3):788–791.

Genewein, T., Leibfried, F., Grau-Moya, J., and Braun, D. A. (2015). Bounded rationality, abstraction, and hierarchical decision-making: An information-theoretic optimality principle. *Frontiers in Robotics and AI*, 2:27.

Gibson, E., Futrell, R., Jara-Ettinger, J., Mahowald, K., Bergen, L., Ratnasingam, S., Gibson, M., Piantadosi, S. T., and Conway, B. R. (2017). Color naming across languages reflects color use. *Proceedings of the National Academy of Sciences*, 114(40):10785–10790.

Gilbert, A., Regier, T., Kay, P., and Ivry, R. (2006). Whorf hypothesis is supported in the right visual field but not the left. *Proceedings of the National Academy of Sciences*, 103:489–494.

Griffin, L. D. (2006). Optimality of the basic colour categories for classification. *Journal of The Royal Society Interface*, 3(6):71–85.

Harremoës, P. and Tishby, N. (2007). The Information Bottleneck revisited or how to choose a good distortion measure. In *IEEE International Symposium on Information Theory*, pages 566–571.

Hendley, C. D. and Hecht, S. (1949). The colors of natural objects and terrains, and their relation to visual color deficiency. *Journal of the Optical Society of America*, 39(10):870–873.

Jameson, K. and D'Andrade, R. G. (1997). It's not really red, green, yellow, blue: An inquiry into perceptual color space. In Hardin, C. L. and Maffi, L., editors, *Color Categories in Thought and Language*, pages 295–319. Cambridge University Press.

Jaynes, E. T. (1982). On the rationale of maximum-entropy methods. *Proceedings of the IEEE*, 70(9):939–952.

Jraissati, Y. and Douven, I. (2018). Delving deeper into color space. *i-Perception*, 9(4).

Kay, P., Berlin, B., Maffi, L., Merrifield, W. R., and Cook, R. (2009). *The World Color Survey*. Stanford: Center for the Study of Language and Information.

Kay, P. and Kempton, W. (1984). What is the Sapir-Whorf hypothesis? *American Anthropologist*, 86:65–79.

Kay, P. and McDaniel, C. K. (1978). The linguistic significance of the meanings of basic color terms. *Language*, 54:610–646.

Kemp, C. and Regier, T. (2012). Kinship categories across languages reflect general communicative principles. *Science*, 336(6084):1049–1054.

Kemp, C., Xu, Y., and Regier, T. (2018). Semantic typology and efficient communication. *Annual Review of Linguistics*, 4(1).

Komarova, N. L., Jameson, K. A., and Narens, L. (2007). Evolutionary models of color categorization based on discrimination. *Journal of Mathematical Psychology*, 51(6):359 – 382.

Kuschel, R. and Monberg, T. (1974). 'We don't talk much about colour here': A study of colour semantics on Bellona Island. *Man, (New Series)*, 9(2):213–242.

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., and Zitnick, C. L. (2014). Microsoft COCO: Common Objects in Context. In Fleet, D., Pajdla, T., Schiele, B., and Tuytelaars, T., editors, *European Conference on Computer Vision (ECCV)*, pages 740–755, Cham. Springer International Publishing.

Lindsey, D. T. and Brown, A. M. (2006). Universality of color names. *Proceedings of the National Academy of Sciences*, 103(44):16608–16613.

Lindsey, D. T. and Brown, A. M. (2014). The color lexicon of American English. *Journal of Vision*, 14(2):17.

Lindsey, D. T., Brown, A. M., Brainard, D. H., and Apicella, C. L. (2015). Hunter-gatherer color naming provides new insight into the evolution of color terms. *Current Biology*, 25(18):2441–2446.

Loreto, V., Mukherjee, A., and Tria, F. (2012). On the origin of the hierarchy of color names. *Proceedings of the National Academy of Sciences*, 109(18):6819–6824.

Marzen, S. E. and DeDeo, S. (2017). The evolution of lossy compression. *Journal of The Royal Society Interface*, 14(130).

McDermott, K. C. and Webster, M. A. (2012). Uniform color spaces and natural image statistics. *JOSA A*, 29(2):A182–A187.

Michel, J.-B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., Team, T. G. B., Pickett, J. P., Hoiberg, D., Clancy, D., Norvig, P., Orwant, J., Pinker, S., Nowak, M. A., and Aiden, E. L. (2011). Quantitative analysis of culture using millions of digitized books. *Science*, 331(6014):176–182.

Regier, T., Carstensen, A., and Kemp, C. (2016). Languages support efficient communication about the environment: Words for snow revisited. *PLoS ONE*, 11(4):e0151138.

Regier, T., Kay, P., and Khetarpal, N. (2007). Color naming reflects optimal partitions of color space. *Proceedings of the National Academy of Sciences*, 104(4):1436–1441.

Regier, T., Kemp, C., and Kay, P. (2015). Word meanings across languages support efficient communication. In MacWhinney, B. and O'Grady, W., editors, *The Handbook of Language Emergence*, pages 237–263. Wiley-Blackwell, Hoboken, NJ.

Roberson, D., Davies, I., and Davidoff, J. (2000). Color categories are not universal: Replications and new evidence from a stone-age culture. *Journal of Experimental Psychology: General*, 129:369–398.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27.

Shannon, C. E. (1959). Coding theorems for a discrete source with a fidelity criterion. *IRE Nat. Conv. Rec*, 4(142-163):1.

Shepard, R. N. (1994). Perceptual-cognitive universals as reflections of the world. *Psychonomic Bulletin & Review*, 1(1):2–28.

Sims, C. R. (2016). Rate–distortion theory and human perception. *Cognition*, 152:181 – 198.

Steels, L. and Belpaeme, T. (2005). Coordinating perceptually grounded categories through language: A case study for colour. *Behavioral and Brain Sciences*, 28(4):469489.

Tishby, N., Pereira, F. C., and Bialek, W. (1999). The Information Bottleneck method. In *Proceedings of the 37th Annual Allerton Conference on Communication, Control and Computing*.

Tkačik, G. and Bialek, W. (2016). Information processing in living systems. *Annual Review of Condensed Matter Physics*, 7(1):89–117.

Webster, M. A., Mizokami, Y., and Webster, S. M. (2007). Seasonal variations in the color statistics of natural images. *Network: Computation in Neural Systems*, 18(3):213–233. PMID: 17926193.

Webster, M. A. and Mollon, J. (1997). Adaptation and the color statistics of natural images. *Vision Research*, 37(23):3283 – 3298.

Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., and Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences*, 104:7780–7785.

Winter, B., Perlman, M., and Majid, A. (2018). Vision dominates in perceptual language: English sensory vocabulary is optimized for usage. *Cognition*, 179:213 – 220.

Xu, Y. and Regier, T. (2014). Numeral systems across languages support efficient communication: From approximate numerosity to recursion. In *Proceedings of the 36th Annual Meeting of the Cognitive Science Society*.

Yendrikhovskij, S. N. (2001). A computational model of colour categorization. *Color Research & Application*, 26:S235–S238.

Yendrikhovskij, S. N., Blommaert, F. J. J., and de Ridder, H. (1998). Optimizing color reproduction of natural images. In *The Sixth Color Imaging Conference: Color Science, Systems, and Applications*, pages 140–145.

Zaslavsky, N., Kemp, C., Regier, T., and Tishby, N. (2018). Efficient compression in color naming and its evolution. *Proceedings of the National Academy of Sciences*, 115(31):7937–7942.

Zaslavsky, N., Kemp, C., Tishby, N., and Regier, T. (2019). Color naming reflects both perceptual structure and communicative need. *Topics in Cognitive Science*, 11(1):207–219.