Coherence of Achromatic, Primary and Basic Classes of Colour Categories

Dimitris Mylonas, Lewis D. Griffin

Dept. Computer of Science, University College London

Corresponding author:	Dimitris Mylonas
email:	d.mylonas@ucl.ac.uk
Address:	66-72 Gower Street
	London WC1E 6EA
	United Kingdom

Keywords

Abstract

A range of explanations have been advanced for the systems of colour names found in different languages. Some explanations give special, fundamental status to a subset of colour categories. We argue that a subset of colour categories, if fundamental, will be *coherent* - meaning that a non-trivial criterion distinguishes them from the other colour categories. We test the coherence of subsets of achromatic (white, black and grey), primary (white, black, red, green, yellow, blue) and basic (primaries plus brown, orange, purple, pink and grey) colour categories in English. Criteria for defining colour categories were expressed in terms of behavioural, linguistic and geometric features derived from colour naming and linguistic usage data; and were discovered using machine learning methods. We find that achromatic and basic colour categories are coherent subsets but not primaries. These results support claims that the basic colour categories have special status, and undermine claims about the fundamental role of primaries in colour naming systems.

1

1   Introduction

Colour naming is the ability to group millions of discriminable colours into a smaller set of categories and denote them by names such as yellow, turquoise and olive green (Pointer & Attridge, 1998; Derefeldt, et al., 2004). These categories are often called linguistic categories because of their correspondence to colour names (Witzel & Gegenfurtner, 2018). A range of neurophysiological (Kay & McDaniel, 1978; Griffin, 2001; Philipona & O'Regan, 2006; Skelton et al., 2017; Witzel & Gegenfurtner, 2018), ecological (Shepard, 1992; Yendrikhovskij, 2001; Mollon & Jordan, 1997) and cultural (Brown & Lenneberg, 1954; Davidoff et al., 1999; Gibson et al., 2017) explanations have been advanced for the underlying mechanisms of colour categories.

Explanations for colour categorization often give a special role to a subset of categories. An early example is Aristotle's (350 B.C.E.) suggestion that five pure colours – crimson, green, cyan, purple and possibly yellow (Sorabji, 1972) – arise from the mixture of white (light) and black (darkness) and from these all the other impure or irregular colours arise as mixtures of the pure. Hering (1878/1964) also appealed to the idea of purity in his proposal that all colours arise from opponent pairs (red versus green, yellow versus blue, and white versus black) of primary colours. Each Hering primary is considered pure in that it contains no quality of the others; and are widely considered an important early perceptual component in the formation of colour categories (Kay & McDaniel, 1978; Kuehni, 2005; Philipona & O'Regan, 2006; Regier, Kay, & Khetarpal, 2007). Other authors (Dimmick, 1925; Boring, 1949) have argued that grey should also be considered an additional achromatic pure colour arising when each of the opponent processes is in a state of equilibrium, but this has been challenged (Quinn, et al., 1985). Berlin & Kay (1969), proceeding from cross-language comparisons, defined Basic Colour Terms (BCTs) as those terms that are a) monolexemic, b) with scope disjoint from any other BCT, c) not restricted to a limited class of objects and d) psychologically salient. In English this criterion identifies the six Hering opponent colours and additionally brown, orange, purple, pink and grey. Here, it is important to distinguish these linguistic colour primaries from the six Hering primaries (his word 'Grundfarben' may be translated in English as 'elemental' colours), which may refer to the three opponent axes of colour sensation. Nevertheless, the relationship between the perceptual and the linguistic primaries is not trivial

as the most typical examples of colour names in different languages correspond roughly to the unique hue settings (Miyahara, 2003; Kuehni, 2005; Regier, Kay & Cook, 2005) and a recent study showed only partial support for whether unique hues are perceptual categories (Witzel & Gegenfurtner, 2018).

In this study, we make the claim that if a subset of colour categories has a foundational role in the system of colour naming then that will be identifiable in their properties and they will be distinguishable from all other colour categories. We formalize this idea as a subset of colour categories forming a *coherent* class, defined by a generalizable membership criterion. We define a criterion to be generalizable if it can be reliably identified from a subset of members of the class. This rules out trivial list-membership style criteria. If we show that some subset of colours *cannot* be distinguished by a generalizable criterion, hence do not form a coherent class, then we suggest that this presents a challenge to any explanation for colour naming that gives that subset a fundamental role, as no evidence of that role exists.

The criteria that we will consider for defining classes of colour categories are defined in terms of features (or attributes) of colour categories. These features are of diverse type. *Linguistic* features relate to the *name* of the category, *behavioural* features relate to the *application* of that name, and *geometric* features relate to the *colour space extent* of the category.

We derive the numerical values of the features from responses to an online colour naming experiment (Mylonas & MacDonald, 2010) and a large dataset of social media posts. For the social media dataset, we used a million random tweets from Twitter posted in English from within Britain. We consider this dataset as representative of ordinary language use. For linguistic features, we use a) name length measured in letters; b) the number of derivative forms (e.g. greener, greenish, and sea green are all consider derivatives of green) in the naming experiment; and c) usage frequency based on counts in the social media dataset. The behavioural features, computed from the naming dataset, are: a) naming frequency, b) response latency and c) inter-subject consensus. The geometric features, computed from the naming dataset, are: the mean colour space location of the distribution of samples that generate the naming response, and the size and shape of that distribution.

Having defined numerical features for a large set of colour categories we are in a position to specify class membership criteria that define subsets of these categories. Although there is a rich and venerable literature on how class membership can be defined such as 'necessary and sufficient conditions' (Berlin & Kay, 1969), 'similarity to prototypes' (Rosch Heider, 1972), or 'networks of family resemblances' (Rosch & Mervis, 1975) these methods have been superseded by techniques developed in the context of computational statistics and machine learning. In particular a 'forest of decision trees' is an extremely general method for expressing class membership. Each decision tree specifies a rule of the form: x is in X if and only if $f_1 < t_1$ or $f_2 > t_2$ where x is an element (here a colour category such as 'red'), X is a class of colour categories (such as the primaries), f_i are features (e.g. naming frequency) describing x, and t_i are thresholds (1 in 10000). In a forest of decision trees each tree is different, and the membership decision of the forest is the majority opinion of the constituent trees.

Effective algorithms ('Random Forests') exist for construction of forests of decision trees based on training examples (e.g. $x_1, \dots x_n$ are in X, $x_{n+1}, \dots, x_N$ are not). The forests that result from these algorithms correctly predict the membership of the training data, and are often very effective at successfully predicting the membership of held-out data not used for training. The key to this generalization success is the technique used to ensure that the trees of the forest are sufficiently diverse. The most important of these are that each tree is constructed from a different random subset of the training data, and the splitting rule at the branch of each tree is not the best possible rule at that branch, but only one *of* the best. Random Forests constructed like this have been shown to be highly effective for many diverse classification problems (Breiman, 2001; Gislason et al., 2006; Cutler et al., 2007). An advantage of them, useful for our application, is that they do not assume commensurate feature dimensions, or normally-distributed features values.

To assess the coherence of a class of colour categories we measure how well it is defined by a *generalizable* criterion. We enforce generalizability by using a *leave-one-out* evaluation: for each colour category (in class or out) we build a random forest classifier using all *other* colour categories, together with their labels as in-class or out; and then evaluate the class membership confidence of the left-out colour category using that classifier. Finally, we evaluate whether the membership confidences of in-class terms are higher than those of out-of-class terms.

In the main paper, we report the coherence of the Hering primary class (black, white, red, green, blue & yellow) and the Berlin & Kay's basic class (Hering's primaries plus purple, orange, pink, grey & brown). Additionally, we report the coherence of an achromatic class (black, grey & white) to check whether smaller classes are necessarily less coherent because they have fewer examples from which to determine a membership criterion. In the Supplementary Material, we report results for other plausible sets of primary and basic colour categories.

2 Methods

2.1 Online Colour Naming Experiment

An online colour naming experiment was designed to collect unconstrained names for presented colour samples (Mylonas & MacDonald, 2010). Participation was voluntary and anonymous and the experimental sessions were conducted after obtaining informed consent (Varnhagen et al., 2005). Colour stimuli were presented sequentially as rectangles (subtending a visual angle of about 3 degrees at a viewing distance of 50 cm) against a neutral grey background with a black outline of 1 pixel. In response to each stimulus, subjects typed any colour descriptor, either a single or multiple words, without time constrain. Typed responses, along with the typing onset delay were recorded. Each subject viewed 20 colour samples randomly from six hundred colour samples in total from the Munsell Renotation Data set (Newhall et al., 1943), including eleven achromatic samples. The colour samples were specified in the sRGB standard colour space for the Internet. To achieve an approximately uniform sampling within the Munsell colour solid, we followed the suggestions of Billmeyer in Sturges and Whitfield (1995). Specifically, a variable number of hues were sampled at different levels of Value and Chroma. At Chroma 2 - ten hues were sampled, whilst at each successive Chroma step the sampled hues were increased by ten. That means from Chroma 8 to the boundaries of the sRGB gamut, all 40 hues were sampled (Mylonas & MacDonald, 2010).

In this study, we consider 10,000 raw responses from 500 British English participants. Typographic conventions (hyphens, commas, parentheses) were replaced with spaces, leading/trailing spaces were removed, and all multi-character spaces were reduced to single spaces. Different word orders (i.e. orange-red or red-orange) were considered as different names. Capitalization was ignored. Common spelling errors (e.g. 'turqose' for 'turquoise')

5

were corrected with supervision. We excluded disruptive observations (1%) including incomplete, numerical and responses written with non-English characters, and responses from participants with possible colour deficiencies (9.7%). This filtering resulted in a dataset for 447 respondents. Their mean age was 33 years ($SD = 13.5$ years). Females provided 63% of the responses while males provided 37%.

The 8940 filtered colour naming responses from the online experiment consist of 1490 distinct names. Many of these were produced infrequently and can be considered rare and idiosyncratic. We restrict our analysis to 73 colour names in wide cultural use which were produced at least 20 times in our data to give us confidence in their measures. This accounts for 62% of the responses. We have confirmed that considering all colour names ($n = 478$) given by more than one observer does not change our main conclusions but produces untrustworthy or empty measurements (e.g. linguistic frequency, median response latency and volume) for uncommon colour names and restricts the visualization of the ranks of all test colour names.

Although online experimentation introduces variability in the stimulus and viewing conditions, we have previously argued that the advantages of a very large subject pool, plus familiarity of the setting for each subject, compensate for that (Mylonas & MacDonald, 2010; Mylonas & MacDonald, 2016; Paramei et al., 2018). In addition, a direct evaluation of the web-based experiment against a laboratory-based experiment produced a better correspondence between the loci of their colour terms than the agreement between previous laboratory-based studies (Mylonas, Griffin & Stockman, 2019; Boynton & Olson, 1987; Sturges & Whitfield, 1995). In Figure S6, we also compare the location of primary colour terms in colour space against the results of a previous study conducted in controlled viewing conditions to show their good agreement (Sturges & Whitfield, 1995).

2.2 Twitter

To examine the frequency of usage of colour names in everyday online conversations, we counted their rates of occurrence in 1,036,103 random tweets downloaded using the Twitter API. Similarly, to the online colour naming experiment, messages in Twitter are given voluntarily and provide greater volume and variability than other refined sources (Corbett & Davies, 1997). We filtered Twitter's public stream with the geo-location coordinates [-5.4,

50.1, 1.7, 55.8] that correspond to a rectangle approximating the extent of the British mainland. We excluded tweets in other languages than English. Each tweet was tokenized using the Natural Language Toolkit (Bird et al., 2009) producing 129,355,280 tokens. Again, typographic conventions and leading/trailing spaces were removed; hyphenated and comma separated and words in parenthesis were treated as multiword colour expressions.

3 Features for colour categories

For each of the 73 common colour names, three sets of behavioural, geometric and linguistic features were computed.

3.1 Behavioural Features

The behavioural features include naming frequency, consensus and response time.

3.1.1 Naming Frequency

Frequency in colour naming experiments quantifies how often each colour name was used to describe any colour stimuli by any observer (Boynton & Olson, 1987; Sturges & Whitfield, 1995). This naming frequency is thus affected by the number of colour samples that evoke that response, and the regularity that they do so. Purple was the most frequent colour name followed by pink, blue and green (Figure S1). The difference between green and the fifth most frequent term - brown - was more than 2% in absolute terms, and more than 50% in relative terms. The least frequent basic term was white, found in the 20th position, while the non-basics lilac and turquoise were found in the 6th and 7th positions respectively.

3.1.2 Consensus

Consensus describes the agreement among observers in naming colour samples (Brown & Lenneberg, 1954; Boynton & Olson, 1987; 1990; Davies & Corbett, 1994; Sturges & Whitfield, 1995). Previous studies have used thresholds for a colour sample being named with consensus, but this approach gives undefined results for rarely named colours. To provide a measure for all colours, in this study consensus is computed by calculating for each colour sample what fraction of the responses are the name, and averaging this fraction over samples where at least one response was the name. Yellow was the colour with the most consistent responses, with orange and pink ranked in the second and third positions respectively (Figure S2). The top 10 ranked colour names were all basic colour terms but grey was ranked in the 13th position following khaki and turquoise.

3.1.3 Reaction time

Reaction time, also called latency, is a measure of the time required to name a colour (Brown & Lenneberg, 1954; Boynton & Olson, 1987; Sturges & Whitfield, 1995). In the online colour naming experiment latencies were measured from the onset of the stimulus to the observer's first keystroke of the typed colour name. Response time distributions are rightwards skewed, so we express their central tendency using the *median* response latency for each colour name rather than the *mean* (Whelan, 2008). White and red were the fastest to name colours and all 11 basic colours were ranked in the top 11 positions (Figure S3).

3.2 Geometric features

The geometric features include the size (volume), shape (anisotropy) and location (centroids) of colour categories in colour space.

3.2.1 Volume (size)

The size of colour categories was measured by their volume in colour space. To approximate the CIELAB volume of the category corresponding to a colour we first described the dispersion of the sample locations evoking each response matching the name of the colour by their covariance matrix. Volume was then measured as the square root of the determinant of this matrix (i.e. the volume of the approximating ellipsoid). To avoid effects of the sampling of colours used in the experiment, that could in principle produce near-to-zero volumes for distributions thin in one direction despite having substantial spread in other directions, we added to the covariance matrix an identity matrix multiplied by the mean colour difference of the four nearest neighbours across stimuli (mean $\Delta E_{ab} = 7.14$). With this way of computing volume, green was the largest category followed by violet and blue (Figure S4). Several basic colours (i.e. white, black and yellow) were not amongst the largest categories in colour space.

3.2.2 Anisotropy (shape)

The shape of colour categories (Gärdenfors, 2004; Jäger, 2010, 2012) was measured by their sphericity assessed from the same covariance matrix used to compute category volume. The logic behind this measure is that if members of the primary class are centred on some 'bumps' of saturation on the uneven surface of the colour solid while secondaries are located in the intermediate regions (Jameson & D'Andrade, 1997; Regier et al., 2007) then this difference will be manifested in their shape and primary categories would be more spherical than non-primaries. The specific quantification used for measuring sphericity was Fractional Anisotropy

(Basser & Pierpaoli, 1996), a size invariant, pure-shape measure that ranges from zero for spherically distributed 3-D data, up to unity for data constrained to a line, hence maximally anisotropic. Intermediate values indicate degrees of anisotropy. Hence, colour categories that are near spherical, whether large or small, will have low anisotropy scores; elongated categories will have high scores; and flattened categories will have intermediate scores (Figure S5). Terracotta was the most and bright blue was the least spherical colour categories. Blue and white were the only basic colours found in the top 10 positions while brown was found in the 3rd last position.

3.2.3 Centroids (location)

Centroids mark the centre of a colour category in colour space. For each colour name, we determined the centroid of the locations giving rise to each matching response using the CIELAB Cartesian-style coordinate system. The Cartesian coordinates $a*$ and $b*$ of the centroid were then converted to the perceptual polar coordinates of Chroma ($C*$) and hue (h). Lime green and fuchsia were the colour names with the highest Chroma (Figure S6) while white and black had respectively the highest and lowest lightness. The comparison of the location of the six centroids of primary terms between a previous study (Sturges & Whitfield, 1995) conducted in controlled viewing conditions and of this study produced a satisfactory agreement with a mean colour difference of $\Delta E_{00} = 5.97$; $STD = 2.88$.

3.3 Linguistic features

The linguistic features include the frequency in ordinary communication, the length of the words and the number of derivative forms.

3.3.1 Linguistic frequency

Linguistic frequency measures the usage of a colour name (Hays et al., 1972). To determine the frequency of colour names in everyday online conversations, we measured their probability of occurrence in 1,036,103 random tweets compiled using the Twitter API. Black followed by white and red were the most frequent colour names in Twitter (Figure S7). The 11 basic colours were found in the top 12 positions. The non-basic term cream ranked in the 4th position but possibly because it also has a common non-colour usage.

3.3.2 Name length

We quantify colour name length by the total number of letters in all words (Brown & Lenneberg, 1954; Berlin & Kay, 1969). This measure correlates with phonetic length, and

across language negatively correlates with frequency of usage (Zipf, 1935; Piantadosi et al. 2011). Basic red and non-basic tan were the colours with the shortest name length (Figure S8). Purple, yellow and orange were the basic colours with the longest name length.

3.3.3 Number of derivative forms

Derivative production is a measure of the number of derivative types of a colour name in colour naming responses (Corbett & Davies, 1997). Specifically, Y is a derivative form of term X if Y contains X as a substring. This definition captures suffixes such as –ish (e.g. greenish) and –er (e.g. greener), and compound colour words such as light green or sea green. Green was found with the largest number of derivative forms followed by blue and pink (Figure S9). Turquoise and lilac were the non-basic terms with largest derivative production in the 10th and 12th position. Black was the basic colour with the smallest number of derivative forms. Note that although we counted derivative forms for this feature, we did not combine their responses together for computation of other features.

3.4 Classifier

We constructed criteria for demarcating classes of colour categories using the Random Forests algorithm (Breiman, 2001). As input the algorithm receives a training dataset of colour categories, each described by a vector of feature values, and associated with a binary label indicating whether it is in-class or out-. Based on this input, the algorithm creates an ensemble (forest) of 100 independently-generated decision-trees. We have confirmed that a larger forest does not change the results.

Each tree is grown using a separate dataset created by bootstrap sampling-with-replacement from the training data. Trees are grown down from a root node at which all training data arrives. At each node a feature dimension is chosen to be the basis for a splitting rule. The choice of dimension is made from a subset of all feature dimensions, chosen randomly for that node. Following the standard recommendation, if there are n feature dimensions, then the subset size is (rounded) square root of n; so in our trees, at each node, a subset of three feature dimensions were considered out of the full eleven. Given the feature subset, the particular feature and threshold value that best segregates the data arriving at the node according to its labels are identified. The arriving data is then sent to left and right sub-nodes according to this criterion. Sub-nodes are iteratively constructed below nodes until leaf nodes are reached that receive only

a single training data sample. After tree construction, the unique dataset generating the tree is discarded but the structure of the tree, the splitting dimension and threshold at each node, and the label of the datum in each leaf node is retained.

After construction of the forest a new datum is classified by passing it through the structure of each tree, directing it to sub-nodes according to its feature values, and recording the label of the leaf node at which it finally arrives. The proportion of trees of the forest that classify it as in-class is the overall in-class classification confidence of the forest.

3.5   Coherence of classes of colour categories

We assess the class coherence of three subsets of colour categories:

Achromatic (n_{in}=3): white, black and grey.

Primary (n_{in}=6): white, black, red, green, yellow and blue.

Basic (n_{in} =11): white, black, red, green, yellow, blue, brown, orange, purple, grey and pink.

For each class, the other colour categories of the common set (n_{out}=73-n_{in}) were considered out-of-class.

3.6   Evaluation of classification

Evaluating classifiers on data on which they were trained is generally misleading. To avoid this, and to ensure that the computed class criteria are generalizable we employ a leave-one-out cross-validation strategy. For each class that we assess we build 73 separate classifiers. Each is trained on 72 colour names, with a different colour name left out. The in-class confidence of each colour name is then computed by the classifier which was trained with it left out. To assess the coherence of a class we quantify the extent to which the class confidences of the in-class colours are higher than those of the out-of-class. For this quantification we use a measure based on rank precision. Precision is the fraction of correct positive classifications to a test class over all positive classifications. MAP is the mean average precision of the ranks at the top k positions, where k is the size of the test class (Voorhees & Harman, 2005). MAP will be 1.00 if all in-class confidences are higher than all out-of-class; 0.00 if all in-class confidences are lower than all out-of-class; and intermediate if the range of in-class confidences overlaps the range of out-of-class.

To examine the importance of features for the coherence of each class, we repeat the full leave-one-out assessment and MAP computation, but with classifiers trained with only a subset of features. The subsets we assessed were: all features except one, two out of three families of features, and single families of features. The importance of features or families of features for each class of colours is quantified by how much the MAP score decreases compared to using all features.

4   Results

Having established three families of features, computed the feature values for each colour category, and determined a procedure for accessing the coherence of a class of colour categories, in this section we examine the cohesion of achromatic, primary and basic classes, and determine the contribution of different features to that coherence.

4.1   Coherence of achromatic class

In our first assessment, we examined the coherence of an achromatic class consisting of black, white and grey. The Random Forests classifier gave all three in-class colours higher confidences than all non-class colour categories, giving a maximum possible MAP score of 1.00. In Figure 1, we present the confidence for each colour category to belong to the achromatic class. We remind the reader that the in-class confidence of each colour category is assessed by a classifier that is trained on all colour categories apart from it. White was the colour category with the highest confidence, followed closely by black. Grey was found in the third position but with lower confidence. Light grey was the out-of-class colour category with the highest in-class confidence.
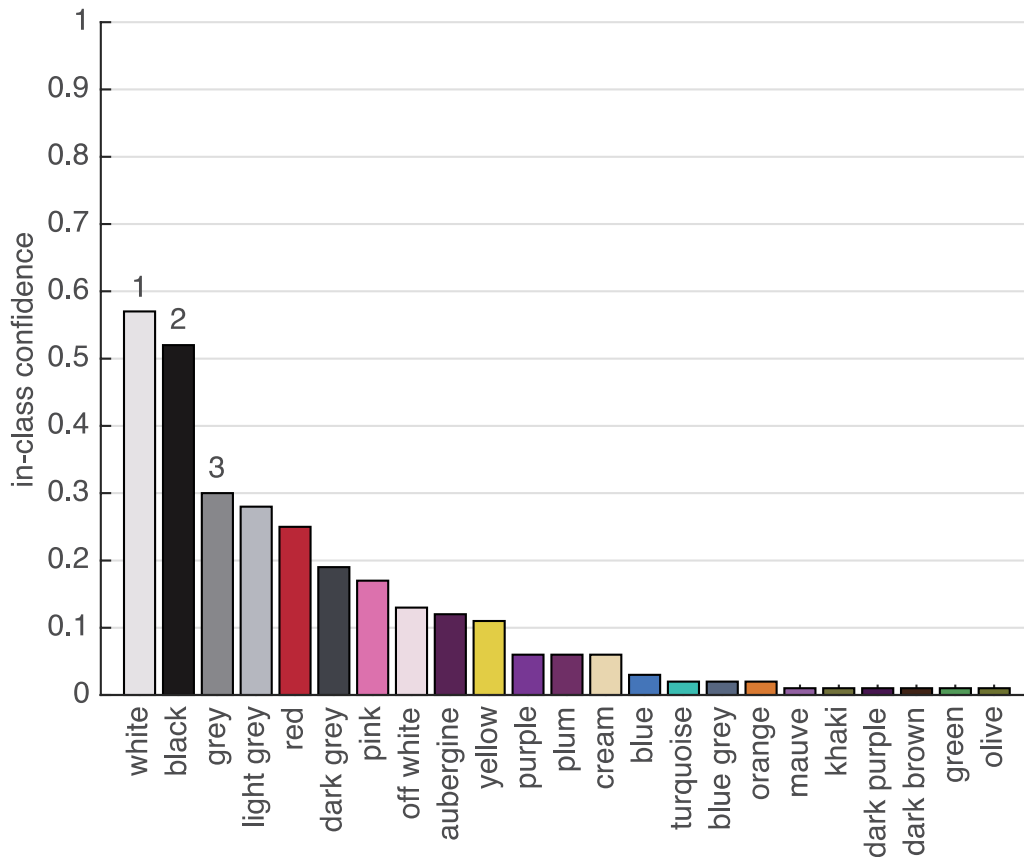
Figure 1. Class confidences for the achromatic class. Only colour categories with an in-class confidence above 0 are shown for clarity. Numbered bars indicate class members. The display colour of each bar is the sRGB centroid of the distribution of samples generating that colour name response.

4.2   Coherence of primary class

As a primary class we took the six suggested linguistic primaries: white, black, red, green, yellow and blue (Berlin & Kay, 1969; Kuehni, 2005; Regier, et al., 2005). The classifier produced a MAP score of 0.50. Examination of the confidences for individual colour categories (Figure 2) showed that this low coherence score was due to failure of the class criteria to generalize to all in-class members (especially yellow), and erroneous generalization to non-class members (especially pink, grey and brown).
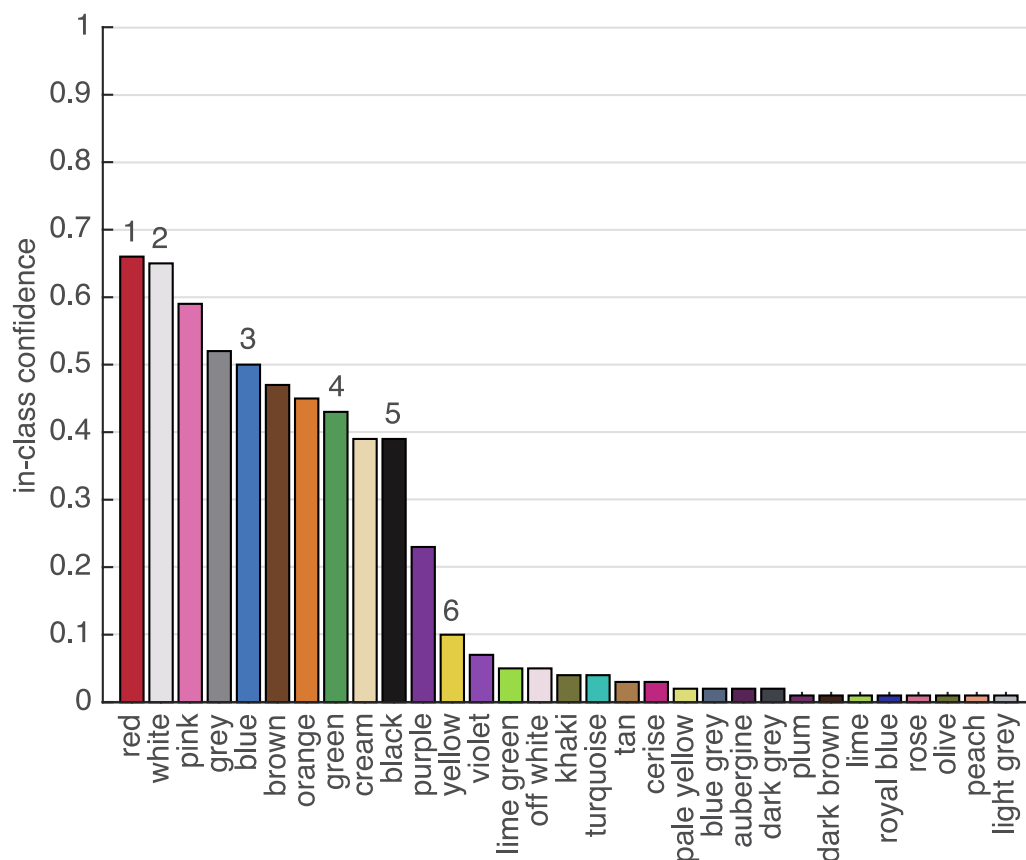
Figure 2. Class confidences for the primary class. See Figure 1 for additional notes.

4.3   Coherence of basic class

For the assessment of the basic class, we considered the 11 basic colour terms of Berlin & Kay (1969), white, black, grey, red, orange, yellow, green, blue, purple, brown and pink. All basic colour categories were given higher confidences than all non-basic, resulting in a maximum possible MAP score of 1.00. Amongst the basics, blue, pink and brown were given the highest confidences and purple the lowest (Figure 3). Amongst the non-basics, olive was given the highest confidence.
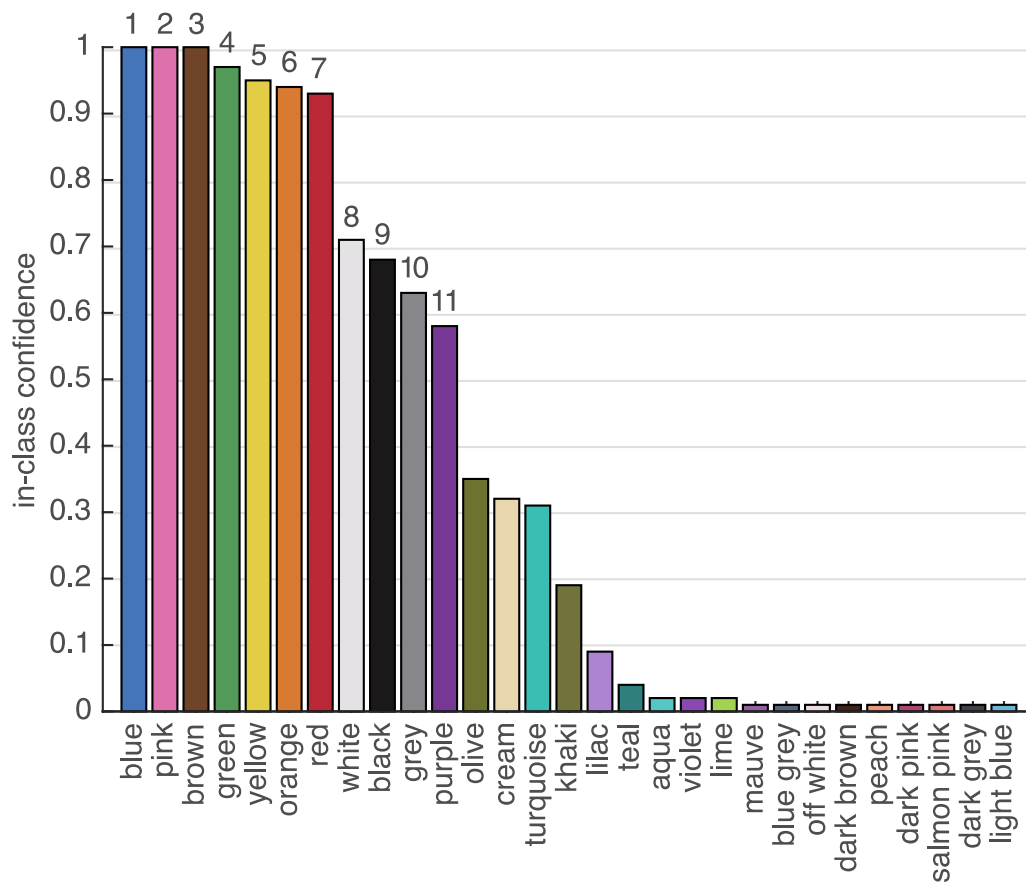
Figure 3. Class confidences for the basic class. See Figure 1 for additional notes.

A summary of all evaluations is given in Table 1.

Table 1. MAP scores, expressing class cohesion, for achromatic, primary and basic classes. A score of 1 is perfect cohesion according to our assessment.

Class	MAP
Achromatic class (n=3)	1.00
Primary class (n=6)	0.50
Basic class (n=11)	1.00

4.4   Feature Contribution

To examine the importance of each feature and each family of features we assessed class coherences using different feature subsets, specifically:

a) All features (*n*=11)

b) All features bar one (*n*=10), eleven variants

c) Behavioural plus Geometric features (*n*=8)

d) Geometric plus Linguistic features (*n*=8)

e) Behavioural plus Linguistic features (*n*=6)

f) Geometric features (*n*=5)

g) Behavioural features (*n*=3)

h) Linguistic features (*n*=3)

Figure 4 summarizes the effect of excluding individual features. For the achromatic class, the greatest effect comes, fittingly, from exclusion of the Chroma feature, which reduces the MAP score from 1.00 to 0.33. Exclusion of consensus, shape, lightness and linguistic frequency had no effect for the achromatic class. For the primary class the most important feature was linguistic frequency, which when excluded reduced the MAP from 0.50 to 0.33. Excluding frequency, response time, size, shape or chroma improved the MAP score. This is presumably because these features are useful to demarcate some of the class, but generalize inconsistently. The greatest improvement was when response time was excluded, raising the MAP score from 0.50 to 0.67. In this case, pink and cream remains as false positives at ranks 5 and 6 with class confidences higher than green and yellow. For the basic class of colour categories none of the excluded features reduced the MAP score below 1.00.
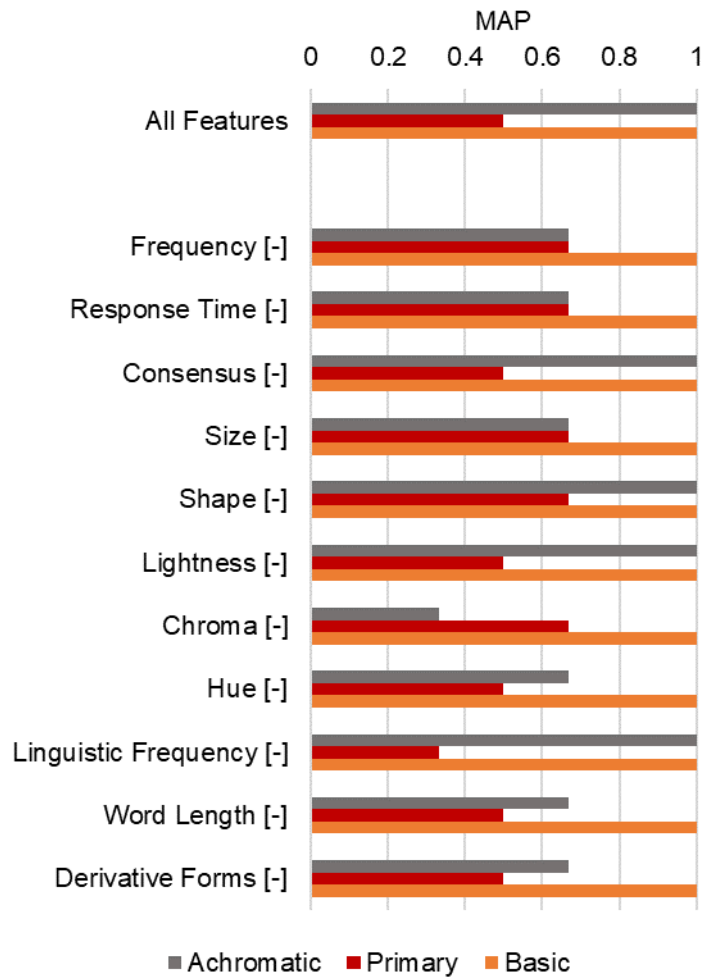
Figure 4. Class coherence scores when a single feature is excluded [-], compared to when all features are used (top).

Considering exclusion of single families of features, for the achromatic class so long as geometric is retained the MAP score is 1.00, otherwise it is 0.33 (Figure 5). For the primary class, the exclusion of linguistic produced the lowest MAP score of 0.33 and the exclusion of behavioural the highest MAP score of 0.66. For basic, the exclusion of geometric and linguistic did not influence the coherence of the class with a MAP score of 1 but excluding the behavioural reduced the MAP score to 0.90 because cream was then given higher confidence than white and black. The assessment of retaining single families of features resulted in a MAP score 0.33 for the achromatic class when either behavioural or linguistic were retained and a MAP score of 1 when geometric was retained. For the primary class, keeping only geometric features produced a MAP score of 0.17 while retaining the linguistic resulted in a maximum

17

MAP score of 0.66. For the basic class, behavioural or linguistic features alone produced a MAP score of 0.90, but geometric alone gave a MAP score of 0.54.
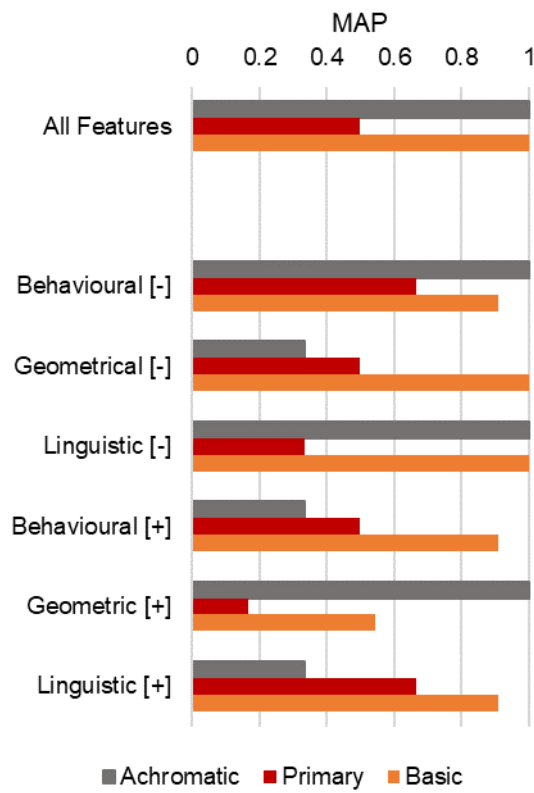

Figure 5. Class coherence scores when one family of features is excluded [-], and when only one family of features is retained [+], compared to all features (top).

5   Discussion

A point of contention that frequently arises regarding the basis of colour categorization is whether there are subsets of colour categories with a special fundamental status. Different subsets have been suggested as fundamental, and no consistent assessment of each of their claims has been previously been made. Here, we argue that a fundamental subset of colour categories should form a coherent class, with a generalizable membership criterion demarcating it. To test this, we analysed large datasets of colour naming responses from an online colour naming experiment and public social media posts to examine the class coherence of achromatic, primary and basic colours. Our findings provide evidence to substantiate the coherence of basic and achromatic classes but we found less support for the primary class. Indeed, the best generalizable criteria for demarcating the primaries consistently also capture

18

secondary colours. These results argue against a set of primary colour categories playing a fundamental role in the wider colour naming systems.

In our assessment of the primary class, we considered the linguistic primaries related to Hering's opponent process theory because of a widely held view that these colour categories are the basis of colour naming systems across languages (Kuehni, 2005; Regier et al., 2005; Philipona & O'Regan, 2006). Still, the number and the members of the primary class vary in the literature (Aristotle, 350 B.C.E.; Newton, 1730; Maxwell, 1872; Hering, 1878/1964; Eskew, 2009; Skelton et al., 2017). In the supplementary section, we tested whether primary classes with different proposed members than those of Hering will be distinguishable from all other colour categories but again we found no evidence to substantiate the coherence of any primary class. The coherence of primary classes proposed by Aristotle (0.57; Figure S10) and Newton (0.57; Figure S11) and Eskew (0.63; Figure S13) were higher than Hering's primary class but those proposed by Maxwell (0.30; Figure S12) and Skelton et al. (0.40; Figure S14) were lower. All these primary classes are smaller ($3 \leq n \geq 7$) than the basic class ($n = 11$) but this does not explain their low MAP scores, since the even smaller achromatic class ($n = 3$) had perfect coherence (MAP = 1.00) because its members have distinctive, common characteristics. Random classes with equal number of randomly selected colour categories ($n = 6$) had an average MAP score of 0.13 (Figure S15), whilst an equally sized class of secondary basics colour categories (brown, orange, purple, pink and grey plus one of Hering's primaries) had an average MAP score of 0.53 (Figure S16). An examination for the status of Hering's primaries within the eleven basic colour categories shown in Figure S17 revealed that the coherence of the primary class retained a MAP score of 0.5 with the non-primaries orange, brown and grey ranking again higher than other primary categories. On the whole, these results indicate that primaries are not a completely haphazard class but are not more coherent than classes of secondary colour categories; consistent with previous studies in adults (Boynton & Olson, 1987), in infants (Franklin, et al., 2008) and in monkeys (Zeki, 1980).

Considering why the class coherence was low for all systems of primary evaluated, we note that yellow (considered primary in all schemes) was consistently given low class-confidence. The particular characteristics of yellow that might explain these results is its narrower distribution (see Figure S4) and higher lightness (see Figure S6) than other chromatic members of the primary class. Interestingly, yellow was absent in Aristotle's original text where he

named only six out of the seven pure categories; is missing from the wavelength sensitivities of cells in V4 reported by Zeki (1980); and produced only partial evidence for being a perceptual category (Witzel & Gegenfurtner, 2018). A second reason for the universally low coherence scores for primary classes were the consistently high in-class confidences given to pink and brown (non-primary in all schemes). Pink and brown, similarly to green and blue, were responded very frequently, in a very short period of time and with very good agreement between subjects. It is also interesting to note that pink and brown appear as a symmetrically related pair within the cognitive structure of the basic colour categories determined through analysis of similarity, relative lightness and adjacency (Griffin, 2001), suggesting that the salience of these two categories may have a shared explanation.

Consistent with our findings, doubts about primary colour categories as the origin of colour categorization have been raised on conceptual grounds (van Brakel, 1993; Jameson & D'Andrade, 1997; Ocelák, 2014) and in a reanalysis of the World Color Survey (Kay et al. 2010) by Jameson (2010). Doubts have also been raised on neurobiological accounts about the priority of primary colour categories over non-primaries in cortical regions as the peak wavelength sensitivity of neurons is distributed through the spectrum while some neurons are sensitive to extra-spectral (e.g. purple) and desaturated (e.g. pale pink) colours (Zeki, 1980; Komatsu, 1992; Bohon et al., 2016). This is not to say that Hering's primaries have no special status at some stage of visual processing that has yet to be found (Dimmick & Hubbard, 1939; Larimer, Krantz, & Ciceronem, 1975; Abramov & Gordon, 1994; Valberg, 2001; Wuerger, Atkinson & Cropper, 2005; see Lindsey & Brown, 2019 for a recent review); but even if they do, this does not penetrate to them being special at the cognitive level.

In contrast to the poor coherence of the primaries, the 11 basic colour categories (Berlin & Kay, 1969) had a perfect MAP score of 1.00. The coherence of the basic class was also apparent when the classifier was trained with reduced features: behavioural or linguistic features alone gave score of 0.90, together 1.00, geometric features contribute little. Coherence of the basic class is unsurprising given that they were originally identified according to a criterion based on features similar to the ones we use. Our results are a confirmation that the Berlin and Kay's basic colour categories can be distinguished from other colour categories by such a criterion in English.

Discussion of the basic colour categories is frequently concerned with why these particular colour terms satisfy this criterion, rather than some other colour terms. Different candidate answers have been advanced, placing different emphasis on the role of physiology or natural world properties. On the one hand, Griffin (2001) has shown that the cognitive similarity structure of the 11 basic colour categories has a symmetry which corresponds to a symmetry of the cone response functions. At the other end of the spectrum of explanations is grounding in the statistical regularities found in natural images (Yendrikhovskij, 2001; but see Steels & Belpaeme, 2005, for arguments that the claim is spurious as different colour spaces produce diverging results), or optimal performance at tasks where semantics must be inferred from appearance (Griffin, 2006). Any explanation, whether it lies in the spectrum, must account for the variation in the number of basic colour names across languages; and some authors have questioned whether the same set of basic colour categories is coherent in all cultures, dependent on the communication needs of semantic categories that are locally most important (Davidoff et al., 1999; Gibson et al., 2017). A cross-language extension of the current methodology could shed light on this.

The examination for a possible *additional* 12th basic colour term in the supplementary section showed a slight deterioration of the coherence of the class, except when cream was added which also produced a perfect MAP score of 1.00. The reversal of the confidence ranking of cream and olive, when olive or cream is added to the basic class (compare Figure 3 and Figure S18 and S19) is surprising but explicable. Consider the category of 'flying birds'. What animal is the closest to being in-class by generalizing from the class? Possibly penguins, with emus further behind. But when penguins are grouped with flying birds, then the criterion which demarcate the class from all other animals would change substantially (promoting the importance of feathers perhaps), and emus could become more in-class than penguins. Cream was also suggested as a candidate for a 12th basic colour terms in a previous study (Sturges & Whifield, 1995) but similarly to our findings with much lower scores than the other 11 basic terms. This indicates that the upper limit of the basic class has some fuzziness and new basic terms may arise (Hardin & Maffi, 1997; Mylonas & MacDonald, 2016; Witzel, 2018).

In the collection of our behavioural colour naming data, we extend earlier studies which used only the most saturated colour samples on the surface of the Munsell system (Berlin & Kay, 1969; Kay et al., 2010; Lindsey et al., 2015; Skelton et al., 2017) by sampling also the interior

of the colour solid. Despite the uneven surface of the Munsell system in terms of saturation (Witzel & Franklin. 2014) that is a necessity for sampling typical colours of colour categories at different chroma steps, like red and pink, we found no higher confidence for the primary terms. Consistent with our findings about the lack of advantage of primaries over non-primaries in colour naming are also results restricted to equiluminant hues of fixed saturation and constrained terms (Emery et al., 2017a). A further methodological improvement includes the departure from usual methods which would use a small number of observers and/or the use of only a restricted set of monolexemic terms (Berlin & Kay, 1969; Boynton & Olson, 1987; Sturges & Whitfield, 1995; Lindsey & Brown, 2014). Instead, thousands of volunteers from linguistically and demographically diverse populations named freely a large number of colours online (Moroney, 2003; Mylonas & MacDonald, 2010). We also depart from previous research which used a refined corpus for the linguistic measurements (Corbett & Davies, 1997) by analyzing a big dataset of real-time Twitter messages in a specific geolocation. We argue that extracting behavioural, geometric and linguistic features of colours from large online datasets allows us to generalize our findings to a larger population sample than earlier studies.

Could our results be influenced by our online experimental methodology, the quality of features and absent features? Regarding the uncontrolled colour reproduction of web-based colour naming experiments, the comparisons against results of previous studies conducted in laboratory conditions produced similar centroids for the basic colour terms in English and in different languages (see Figure S8; Moroney, 2003; Mylonas & MacDonald, 2010; Mylonas & MacDonald, 2016; Paramei et al., 2018). An assessment of the precision of our uncalibrated colour naming experiment conducted over the Internet against a calibrated experiment – using the same sample set and background – performed in a laboratory environment (Mylonas, Griffin & Stockman, 2019) showed also superior agreement than the comparisons between previous laboratory-based studies (Boynton & Olson, 1987; Sturges & Whitfield, 1995). Furthermore, the response times reported here, albeit longer than latencies recorded in laboratory settings, replicates the advantage of the basic terms and the equality of primary and secondary basic terms reported in previous studies (Boynton & Olson, 1987, Corbett & Davies, 1997). Collectively these results suggest that crowdsourcing- and laboratory- based colour naming experiments produce consistent results and support the validity of both methods in estimating colour naming functions in laboratory and real-world monitor settings.

With respect to different computational approaches for determining the features of each colour category, we recognize that there are alternative reasonable ways to compute some of these. For example, replacing the reported *median* response time with the *mean* as used in previous studies (Boynton and Olson, 1987) or replacing the probabilistic calculation of consensus of this study with a more information-based computation (Gibson et al., 2017). We have not found that variants of computations for either response time or consensus substantially alters our results.

A possible missing feature could be the purity of each colour category. Purity is related to our naming consensus measure but it could be argued that a hue cancelation task would provide a better measure. Nevertheless, previous studies (Malkoc, Kay & Webster, 2005; Bosten & Boehm, 2014) found no differences between unique-hue judgments of non-primary (i.e., orange, purple) and primary hues (i.e. red, yellow, green and blue), suggesting that inclusion of such a feature would not be sufficient to make the primaries coherent. The lack of advantage for unique hues over non-unique hues has also been reported in visual search and hue scaling tasks (Wool et al., 2015; Emery et al., 2017b).

A different type of missing feature would be relational features, such as the small colour differences between category members and large colour differences between members of different categories, employed in recent computational methods (Regier et al., 2007; Regier, Kemp & Kay, 2015; Zaslavsky et al., 2018). For example, Regier and his colleagues (2007) suggested that there might be optimal ways of dividing the surface of the colour solid into the 6 colour categories of Hering (white, black, red, yellow, green but blue-purple are combined into a single category) based on the uneven shape of perceptual colour space where several large 'bumps' of saturation presumably produce areas with greater consensus among speakers across languages (Jameson & D'Andrade, 1997). Our class coherence approach is not accommodating relational features directly since they belong jointly to a class, not separately for each colour category. However, by sampling the surface and also the interior of the colour solid, we found that neither their shape (measured by Fractional Anisotropy), nor consensus, not even saturation was sufficient to demarcate Hering's primaries from all other categories. Indeed, we consider that the most compelling justification for most systems of primaries is not their fundamental role in colour categorization but their practical success in subtractive or additive colour mixing.

In summary, we show that primary colour categories do not form a coherent class, whilst achromatic and basic classes do. These results provide evidence against primaries playing a fundamental role in the development of colour naming systems and support the particular role of basic colour categories.

Acknowledgments

References

Abramov, I., & Gordon, J. (1994). Color Appearance: on Seeing Red-or Yellow, or Green, or Blue. *Annual Review of Psychology*, *45*(1), 451–485.

Aristotle. (350 B.C.E.). *On Sense and the Sensible*. (J. I. Beare, Trans.). Adelaide: The University of Adelaide Library.

Basser, P. J., & Pierpaoli, C. (2011). Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor MRI. *Journal of Magnetic Resonance*, *213*(2), 560–570.

Berlin, B., & Kay, P. (1969). *Basic color terms: their universality and evolution*. Stanford, Calif.: Center for the Study of Language and Information.

Bird, S., Klein, E., & Loper, E. (2009). *Natural Language Processing with Python* (1 edition). Beijing; Cambridge Mass.: O'Reilly Media.

Bohon, K. S., Hermann, K. L., Hansen, T., & Conway, B. R. (2016). Representation of Perceptual Color Space in Macaque Posterior Inferior Temporal Cortex (the V4 Complex). *ENeuro*, *3*(4).

Boring, E. G. (1949). I. - A color solid in four dimensions. *L'année Psychologique*, *50*(1), 293–304.

Bosten, J. M., & Boehm, A. E. (2014). Empirical evidence for unique hues? *Journal of the Optical Society of America A*, *31*(4), A385.

Boynton, R. M., & Olson, C. X. (1987). Locating basic colors in the OSA space. *Color Research & Application*, *12*(2), 94–105.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32.

Brown, R. W., & Lenneberg, E. H. (1954). A study in language and cognition. *The Journal of Abnormal and Social Psychology*, *49*(3), 454–462.

Corbett, G. G., & Davies, I. R. L. (1997). Establishing Basic Colour Terms: Measures and Techniques. In C. L. Hardin & L. Maffi (Eds.), *Color Categories in Thought and Language* (pp. 197–223). Cambridge: Cambridge University Press.

Cutler, D. R., Edwards, T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., & Lawler, J. J. (2007). Random Forests for Classification in Ecology. *Ecology*, *88*(11), 2783–2792.

Davies, I. L., & Corbett, G. G. (1995). A practical field method for identifying probable basic colour terms. *Languages of the World*, *9*(1), 25–36.

Davidoff, J., Davies, I., & Roberson, D. (1999). Colour categories in a stone-age tribe. *Nature*, *398*(6724), 203–204.

Derefeldt, G., Swartling, T., Berggrund, U., & Bodrogi, P. (2004). Cognitive color. *Color Research & Application*, *29*(1), 7–19.

Dimmick, F. L. (1925). Discussion: The Series of Blacks, Grays, and Whites. *Psychological Review*, *32*(4), 334–336.

Dimmick, F. L., & Hubbard, M. R. (1939). The Spectral Location of Psychologically Unique Yellow, Green, and Blue. *The American Journal of Psychology*, *52*(2), 242–254. JSTOR. https://doi.org/10.2307/1416110

Emery, K. J., Volbrecht, V. J., Peterzell, D. H., & Webster, M. A. (2017a). Variations in normal color vision. VII. Relationships between color naming and hue scaling. *Vision Research*, *141*, 66–75. https://doi.org/10.1016/j.visres.2016.12.007

Emery, K. J., Volbrecht, V. J., Peterzell, D. H., & Webster, M. A. (2017b). Variations in normal color vision. VI. Factors underlying individual differences in hue scaling and their implications for models of color appearance. *Vision Research*, *141*, 51–65. https://doi.org/10.1016/j.visres.2016.12.006

Eskew Jr., R. T. (2009). Higher order color mechanisms: A critical review. *Vision Research*, *49*(22), 2686–2704.

Franklin, A., Pitchford, N., Hart, L., Davies, I. R. L., Clausse, S., & Jennings, S. (2008). Salience of primary and secondary colours in infancy. *British Journal of Developmental Psychology*, *26*(4), 471–483.

Gärdenfors, P. (2004). *Conceptual Spaces: The Geometry of Thought*. Cambridge, Mass.: MIT Press.

Gibson, E., Futrell, R., Jara-Ettinger, J., Mahowald, K., Bergen, L., Ratnasingam, S., … Conway, B. R. (2017). Color naming across languages reflects color use. *Proceedings of the National Academy of Sciences*, 201619666.

Gislason, P. O., Benediktsson, J. A., & Sveinsson, J. R. (2006). Random Forests for land cover classification. *Pattern Recognition Letters*, *27*(4), 294–300.

Griffin, L. D. (2001). Similarity of psychological and physical colour space shown by symmetry analysis. *Color Research & Application*, *26*(2), 151–157.

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

Hardin, C. L., & Maffi, L. (1997). Introduction. In *Color Categories in Thought and Language*. Cambridge University Press.

Hays, D. G., Margolis, E., Naroll, R., & Perkins, D. R. (1972). Color Term Salience. *American Anthropologist*, *74*(5), 1107–1121.

Hering, E. (1878/1964). *Outlines of a theory of the light sense*. (L. M. Hurvich & D. Jameson, Trans.). Cambridge, Mass.: Harvard University Press.

Jäger, G. (2010). Natural Color Categories Are Convex Sets. In M. Aloni, H. Bastiaanse, T. de Jager, & K. Schulz (Eds.), *Logic, Language and Meaning* (Vol. 6042, pp. 11–20). Springer Berlin Heidelberg. http://link.springer.com/10.1007/978-3-642-14287-1_2

Jäger, G. (2012). Using Statistics for Cross-linguistic Semantics: A Quantitative Investigation of the Typology of Colour Naming Systems. *Journal of Semantics*. https://doi.org/10.1093/jos/ffs006

Jameson, K., & D'Andrade, R. (1997). It's not really red, green, yellow, blue: an inquiry into perceptual color space. In C. L. Hardin & L. Maffi (Eds.), *Color Categories in Thought and Language* (pp. 197–223). Cambridge: Cambridge University Press.

Jameson. K. A. (2010). Where in the World Color survey is the support for the Hering primaries as the basis for color categorization? In J.D. Cohena and M. Matthen (Eds.) *Color Ontology and Color Science,* (pp. 179-202). Cambridge Mass: MIT Press.

Kay, P., & McDaniel, C. K. (1978). The Linguistic Significance of the Meanings of Basic Color Terms. *Language*, *54*(3), 610–646.

Kay, P., Berlin, B., Maffi, L., Merrifield, W., & Cook, R. (2010). *The World Color Survey*. Berkley, CA: CSLI Publications.

Komatsu, H., Ideura, Y., Kaji, S., & Yamane, S. (1992). Color selectivity of neurons in the inferior temporal cortex of the awake macaque monkey. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *12*(2), 408–424.

Kuehni, R. G. (2005). Focal Color Variability and Unique Hue Stimulus Variability. *Journal of Cognition and Culture*, *5*(3), 409–426.

Larimer, J., Krantz, D. H., & Cicerone, C. M. (1975). Opponent process additivity—II. Yellow/blue equilibria and nonlinear models. *Vision Research*, *15*(6), 723–731. https://doi.org/10.1016/0042-6989(75)90291-6

Lindsey, D. T., & 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., & Apicella, C. L. (2015). Hunter-Gatherer Color Naming Provides New Insight into the Evolution of Color Terms. *Current Biology*, *25*(18), 2441–2446.

Lindsey, D. T., & Brown, A. M. (2019). Recent progress in understanding the origins of color universals in language. *Current Opinion in Behavioral Sciences*, *30*, 122–129.

McLaren K. (1985). Newton's indigo. *Color Research & Application*, *10*(4), 225–229.

Malkoc, G., Kay, P., & Webster, M. A. (2005). Variations in normal color vision. IV. Binary hues and hue scaling. *Journal of the Optical Society of America A*, *22*(10), 2154.

Maxwell, J. C. (1872). On colour vision. In *Roy. Instit. Proc. VI. Scientific Papers* (Vol. 1, pp. 260–271).

Mollon, J. D., & Jordan, G. (1997). On the Nature of Unique Hues. In *John Dalton's Colour Vision Legacy* (pp. 381–392). London: Taylor & Francis.

Moroney, N. (2003). Unconstrained web-based color naming experiment. *Color Imaging: Processing, Hardcopy and Applications VIII (Proceedings of SPIE/IS&T)*, *Volume 5008*, 36–46.

Mylonas, D., & MacDonald, L. (2010). Online Colour Naming Experiment Using Munsell Samples. In *5th European Conference on Colour in Graphics, Imaging, and Vision and 12th International Symposium on Multispectral Colour Science*. Joensuu, Finland: IS&T.

Mylonas, D., & MacDonald, L. (2016). Augmenting basic colour terms in English. *Color Research & Application*, *41*(1), 32–42.

Mylonas, D., Griffin, D. L., & Stockman, A. (2019). *Mapping colour names in cone excitation space*. 25th ICVS, Riga, Latvia.

Newhall, S. M., Nickerson, D., & Judd, D. B. (1943). Final Report of the O.S.A. Subcommittee on the Spacing of the Munsell Colors. *Journal of the Optical Society of America*, *33*(7), 385–411.

Newton, S. I. (1730). *Opticks:: Or, A Treatise of the Reflections, Refractions, Inflections and Colours of Light*. William Innys at the West-End of St. Paul's.

Ocelák, R. (2014). The Myth of Unique Hues. *Topoi*, *34*, 513–522.

Paramei, G. V., Griber, Y. A., & Mylonas, D. (2018). An online color naming experiment in Russian using Munsell color samples. *Color Research & Application*, *43*(3), 358–374.

Philipona, D. L., & O'Regan, J. K. (2006). Color naming, unique hues, and hue cancellation predicted from singularities in reflection properties. *Visual Neuroscience*, *23*(3–4), 331–339.

Piantadosi, S. T., Tily, H., & Gibson, E. (2011). Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences*, *108*(9), 3526–3529.

Pointer, M. R., & Attridge, G. G. (1998). The number of discernible colours. *Color Research & Application*, *23*(1), 52–54.

Quinn, P. C., Wooten, B. R., & Ludman, E. J. (1985). Achromatic color categories. *Perception & Psychophysics*, *37*(3), 198–204.

Regier, T., Kay, P., & Cook, R. S. (2005). Focal colors are universal after all. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(23), 8386–8391.

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

Regier, T., Kemp, C., & Kay, P. (2015). Word Meanings across Languages Support Efficient Communication. In BriancWhinney & W. O'Grady (Eds.), *The Handbook of Language Emergence* (pp. 237–263). John Wiley & Sons, Inc.

Rosch Heider, E. (1972). Probabilities, Sampling, and Ethnographic Method: The Case of Dani Colour Names. *Man*, *7*(3), 448.

Rosch, E., & Mervis, C. B. (1975). Family Resemblances: Studies in the Internal Structure of Categories. *Cognitive Psychology*, *7*(4), 573–605.

Shepard, R. N. (1992). The perceptual organization of colors: An adaptation to regularities of the terrestrial world? In *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 495–532). New York, NY, US: Oxford University Press.

Skelton, A. E., Catchpole, G., Abbott, J. T., Bosten, J. M., & Franklin, A. (2017). Biological origins of color categorization. *Proceedings of the National Academy of Sciences*, *114*(21), 5545–5550.

Sorabji, R. (1972). Aristotle, Mathematics, and Colour. *The Classical Quarterly*, *22*(2), 293–308.

Steels, L., & Belpaeme, T. (2005). Coordinating perceptually grounded categories through language: A case study for colour. *The Behavioral and Brain Sciences*, *28*(4), 469–489; discussion 489-529. https://doi.org/10.1017/S0140525X05000087

Sturges, J., & Whitfield, A. (1995). Locating basic colours in the Munsell Space. *Color Research and Application*, *Volume 20*(Number 6), 364–376.

Valberg, A. (2001). Unique hues: an old problem for a new generation. *Vision Research*, *41*(13), 1645–1657.

Van Brakel, J. (1993). The plasticity of categories: the case of colour. *British Journal for the Philosophy of Science, 24*, 103-135.

Varnhagen, C. K., Gushta, M., Daniels, J., Peters, T. C., Parmar, N., Law, D., Hirsch, R., Takach, B. S., & Johnson, T. (2005). How informed is online informed consent? *Ethics & Behavior*, *15*(1), 37–48.

Voorhees, E. M., & Harman, D. K. (2005). *TREC: Experiment and Evaluation in Information*. The MIT Press.

Whelan, R. (2008). Effective Analysis of Reaction Time Data. *Psychological Record*, *58*(3), 475–482.

Witzel, C., & Franklin, A. (2014). Do focal colors look particularly "colorful"? *Journal of the Optical Society of America A*, *31*(4), A365–A374. https://doi.org/10.1364/JOSAA.31.00A365

Witzel, C., & Gegenfurtner, K. R. (2018). Are red, yellow, green, and blue perceptual categories? *Vision Research*.

Witzel, C. (2018). Misconceptions About Colour Categories. *Review of Philosophy and Psychology*, 1–42.

Wool, L. E., Komban, S. J., Kremkow, J., Jansen, M., Li, X., Alonso, J.-M., & Zaidi, Q. (2015). Salience of unique hues and implications for color theory. *Journal of Vision*, *15*(2), 10–10. https://doi.org/10.1167/15.2.10

Wuerger, S. M., Atkinson, P., & Cropper, S. (2005). The cone inputs to the unique-hue mechanisms. *Vision Research*, *45*(25–26), 3210–3223.

Yendrikhovskij, S. (2001). A computational model of colour categorization. *Color Research & Application*, *26*(S1), S235–S238.

Zaslavsky, N., Kemp, C., Regier, T., & Tishby, N. (2018). Efficient compression in color naming and its evolution. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(31), 7937–7942. https://doi.org/10.1073/pnas.1800521115

Zeki, S. (1980). The representation of colours in the cerebral cortex. *Nature*, *284*(5755), 412–418.

Zipf, G. (1935). *The Psychobiology of Language: An Introduction to Dynamic Philology*. Cambridge, Mass.: M.I.T. Press.