Augmenting Basic Colour Terms in English

Dimitris Mylonas,[1]* Lindsay MacDonald[2]
[1]School of Electronic Engineering and Computer Science, Queen Mary University of London, London E1 4NS, United Kingdom
[2]Civil, Environmental & Geomatic Engineering, University College London, London WC1E 6BT, United Kingdom

Originally published in Color Research and Application Copyright 2015 Wiley Periodicals. The publisher's version is available at:
http://onlinelibrary.wiley.com/doi/10.1002/col.21944/abstract
Please cite the publisher's version.

This document is the author's final manuscript version of the journal article, incorporating any revisions agreed during peer review. Some differences between this version and the publisher's version remain. You are advised to consult the publisher's version if you wish to cite from it.

Augmenting Basic Colour Terms in English

Dimitris Mylonas
School of Electronic Engineering and Computer Science, Queen Mary University of London, London E1 4NS, United Kingdom

Lindsay MacDonald
Civil, Environmental & Geomatic Engineering, University College London, London WC1E 6BT, United Kingdom

Correspondence:
Dimitris Mylonas
email: d.mylonas@qmul.ac.uk
School of Electronic Engineering and Computer Science,
Queen Mary University of London
London E1 4NS, United Kingdom

Abstract
In an unconstrained colour naming experiment conducted over the web, 330 participants named 600 colour samples in English. The 30 most frequent monolexemic colour terms were analyzed with regards to frequency, consensus among genders, response times, consistency of use, denotative volume in the Munsell and OSA colour spaces and inter-experimental agreement. Each of these measures served for ranking colour term salience; rankings were then combined to give a composite index of basicness. The results support the extension of English inventory from the 11 basic colour terms of Berlin and Kay to 13 terms by the addition of lilac and turquoise.

Introduction

The human visual system is able to discriminate millions of different colours,[1] but for practical purposes and everyday communication we tend to organize them into a smaller set of colour categories and give them common names such as ‘red’ or ‘light blue’. Psychophysical colour-naming experiments offer the most direct and legitimate method of determining the mapping between colour names and the corresponding regions of perceptual colour spaces. Over recent years, colour naming data derived from such experiments has been used for image processing,[2–4] computer vision[5,6] and gamut mapping.[7,8]

Colour names are used to signify regions of colour space with empirical significance, and have been found to play an important role in long-term memory and to enhance colour recognition.[9,10] Most languages have a large number of names to describe colours. Like all words, they are subject to fashion and may change their meanings over time. Yet, there exists a small number of basic colour terms (BCTs) that are shared and comprehended well by most speakers in each language. Unconstrained colour naming experiments are able to capture a great deal of the richness of colour language, including single and multiple word descriptions, but establishing the basic terms in each language has proved to be a difficult task that requires multiple criteria and appropriate tests.[11, 12]

The gamut of colours perceived by a normal trichromatic observer is a manifold in a three-dimensional colour space. Each colour from the visual environment can be mapped onto a point within this colour space, and a colour term can be defined by the extent of the applicable region. We are interested in finding the distribution of each colour term, and also the location within each region of the point representing its center-of-mass, called centroid. Because perceived colour space is a continuum, with no intrinsic restrictions on how it can be mapped into a lexicon of colour terms, it would seem that any number of arbitrary mappings would be equally valid. In practice, speakers of diverse languages show a surprising degree of consensus, [13] especially for focal colours, with the inter-language differences being less than the intra-language differences among speakers.[11] It has been hypothesized that languages gravitate to an optimal set of categories and return to them despite departures from the norm by individual speakers.[14–16]

The main question underlying this study is what is meant by ‘basicness’. This goes to the heart of the meaning of categories, and why people seek to classify an object or a perceived characteristic of an object as being in one category or another. The tendency to discriminate one thing from another is inextricably linked with learning and recognition. To categorize a situation as ‘A’ not ‘B’ may mean the difference between safety and danger, even life and death, for example the identification of the colour of the traffic lights on the road ahead.

Dummett[17] noted the distinction between categories of names (linguistic designations) and categories of the entities to which the names refer. He argued that categorical concepts are necessary for us to single out ‘things’ in every situation. So the category differences between the names we use are, ipso facto, also category differences between the things singled out by these names. In this way, the connection between the category of an expression used to refer to a given entity (the ‘reference’) and the category of the entity referred to (the ‘referent’) is ensured.

For communication between people in a society, however, more is needed beyond the categorical association of a name with an entity. Each individual might have his or her own idiosyncratic system of categories, but for meaningful exchange with others there has to be some shared understanding and commonality of categories throughout the social group. At the simplest level, this begins with monosyllabic utterances referring to physical entities in the surrounding environment familiar to every individual, such as earth, sea, sky, snow, etc. These are the basic terms upon which human communication and language are built.

In the case of colour, the referent is not a physical object but a sensation, something experienced by an individual. The colour name is a label, shared by members of a language group, on the assumption that when individuals look at the same object they experience a similar sensation, or at least that the individual sensations they experience can be categorized with a common name. In many cases, colour names arise by familiar association with food.[18] Thus, orange is not only the name of an edible fruit, but also a signifier for the colour of that fruit. By abstraction then orange becomes a metonym for all objects that give rise to the same group of visual sensations.

Basicness in the context of colour names means the minimal set of linguistic signifiers that individuals within a language group can use to communicate their categories of colour sensation. This depends on the responses of the human visual system, the referents in the environment of the social group and the stage of development of the language.

To qualify as a BCT, a colour name in our view should:
a. be widely used in a population of speakers;
b. have a shared meaning for the associated colour stimulus;
c. be salient in the sense that the colour is easily identifiable in an array; and
d. be reliably distinguishable from its neighbours in colour space.

Criterion (d) means that a basic colour name should retain its identity when inverted, i.e. that the centroid of a colour name should not be subsumed by larger neighbour categories.

This article explores the notion of basicness using a large set of empirical responses from an online colour naming experiment[19] and with a set of appropriate behavioural measures. The collected names are analyzed by the frequency of usage of colour terms, consensus among genders, consistency of responses, response times, denotative volume in colour space and inter-experimental agreement. We explore, in particular, whether the number of BCTs in English should be extended beyond the 11 established by Berlin and Kay.[11] This would have useful applications for improving the precision of colour naming in colorimetric colour spaces and for facilitating colour communication within and between different cultures over global networks.[20]

Related Studies

Colour naming research is an interdisciplinary area that brings together colour science, psychology, anthropology, linguistics and computer science. Brown and Lenneberg[21] carried out a colour naming study in support of the linguistic relativity principle, also known as the Saphir–Whorf hypothesis, which suggests that linguistic categories available in a certain language influence cognitive classifications by speakers of that language and, consequently, the way that they think and behave.

In contrast, Berlin and Kay[11] asserted that all languages can have up to 11 universal BCTs constrained by the physiologically grounded human visual system.[22] These undergo a seven-stage evolution in the development of colour vocabulary with the following order of emergence:

Stage I: Black and white
Stage II: Red
Stage III: Either green or yellow
Stage IV: Both green and yellow

Stage V: Blue
Stage VI: Brown
Stage VII: Purple, pink, orange, gray

Berlin and Kay (B&K) defined basicness by a combination of linguistic and psychological criteria:

- A BCT should be monolexemic.
- Its scope should not overlap with any other colour term (e.g., the meaning of navy is a hyponym of blue).
- It should not be restricted to a limited class of objects (e.g., blond describes only hair colour or beer).
- It should be psychologically salient for speakers of the language in question.
- Its meaning is not divisible or determined by its parts (e.g., greenish).

B&K used an elicitation method to identify the most common colour terms in each language and then employed an array of 330 samples (Mercator projection of the Munsell solid) for mapping these terms, identifying the best example (prototype) of each term. The first six BCTs, i.e. white, black, red, green, yellow and blue, are called 'primary basic', while the remaining five terms are called 'derived' or 'secondary basic'.[22] It is important to distinguish these linguistic colour primaries from the six Hering primaries (his word 'Grundfarben' may be translated in English as 'component' or 'elemental' colours), which refer to the three opponent axes of colour sensation.[23]

Subsequent studies substantiated the universal inventory but also revealed variations and differences, even between languages with the same number of BCTs. Boynton and Olson[24] conducted a colour naming experiment with American English to locate the denotata of the BCTs in the OSA space. The experiment involved 424 uniformly spaced colour samples, presented against a neutral gray background of 20% reflectance under a photoflood lamp of 3,200 K. Response times (RTs) were measured from the onset of the stimulus to the start of the subject's vocalization. Observers were asked to use solely monolexemic colour terms. Their study showed that the 11 BCTs were used more frequently, more consistently, with greater consensus and more quickly than non-BCTs. The authors also suggested an emergent twelfth BCT in the region between white, yellow, orange, pink and brown. The word most frequently used for that region was peach but it was not qualified as a BCT.

Following the Boynton and Olson study, Sturges and Whitfield[25] located denotata of the BCTs in the Munsell system for British English. The experiment involved 446 colour samples presented randomly against a neutral gray background of Munsell N7 (matte) under a CIE D65 simulator. Their results confirmed that BCTs have shorter response times and higher consistency and consensus than non-BCTs. An interesting finding was that purple ranked third in terms of consistency and frequency, along with short response times, and appeared to cover a larger area of the Munsell than the OSA space. Cream was suggested as a candidate for a twelfth BCT, as it was used frequently and consistently but with a clear differentiation from the 11 BCTs.

Davies *et al.*[26] proposed a relatively faster method of identifying BCTs in English. The procedure included two tasks: first listing any colour names, and second mapping the names onto a set of colour tiles. The authors estimated saliency of a colour term as a combined index of both frequency of the term in the listing task and consensus in the mapping task. Notably, turquoise was reported as one of the most frequent non-BCTs. Kerttula[27] defined four parameters for basicness: (1) primacy, expressing how primary is the colour sense of the term; (2) frequency, giving the number of occurrences in a text or discourse; (3) application, defining the number of referents and (4) derivational productivity, conveying the number of derivative words or compounds. She developed the concept of 'relative basicness', as the degree to which colour terms are established in relation to each other. In an online experiment with unconstrained colour naming, Moroney[28] used 'distributed psychophysics', as he called it, to collect a small number of colour names from a large number of observers over the web. Participants were asked to give the best names for seven patches of colours selected randomly from a 6 X 6 X 6 non-perceptually uniform grid sampling of the RGB cube, viewed on a desktop display against a white background. Results of the online experiment were validated against the results of Boynton and Olson[24] and Sturges and Whitfield,[25] both obtained under controlled laboratory conditions, and showed a high degree of correlation with the chromatic basic colours, expressed as hue angles in CIELAB.

Currently, a balanced view reconciles both relativist and universalist theories and allows some degree of language specific variation in the cognitive organization of colour.[29–31] The 11 BCTs divide colour space coarsely into the corresponding colour categories[32] and there is no physiological basis for considering all basic terms equivalent.[22] The way is therefore open for languages to acquire more than 11 BCTs, and secondary terms constitute a group of potential candidates for the emergence of new BCTs.[33]

Experiment

We have designed an online multilingual colour naming experiment, accessible at www.colournaming.com, to collect broad datasets of colour names from a large number of observers from linguistically and demographically diverse populations.[19] Over the past seven years, the experiment has been translated into 14 languages and has gathered responses from many thousands of observers.

At the beginning of the experimental procedure, we ask the observer to adjust his/her display to sRGB settings using an advanced or basic set of instructions and the brightness of the monitor to make visible all 21 steps of a gray scale ramp. We also screen the observer for possible colour deficiencies with a web-based Dynamic Colour Vision Test developed at the City University London.[34]

In the unconstrained colour-naming task, each participant is presented with a sequence of 20 colours randomly selected from 600 total samples in the Munsell Renotation Dataset. Following the suggestions of Billmeyer (cf. Ref. 25), the 600 samples were chosen as an approximately uniformly distributed array from a variable number of hues at different Munsell value and chroma. Colour stimuli were specified in sRGB and presented against a neutral mid-gray background. Stimulus size (width by height) on the display was 147 by 94 pixels, which for a display resolution of 3.3

pixels per mm (83 pixels per inch) would be 45 by 30 mm, subtending an angle of approximately 5 by 3.4 degrees at a viewing distance of 50 cm.

Across 330 observers, each colour sample for the dataset used in this study was presented on average 9.04 times (σ=3.04), while one sample was presented twice to each observer to estimate naming consistency. Response times were measured from the onset of the stimulus to the subject's first keystroke of the typed colour name. The web interface also includes two questionnaires to collect information about the viewing conditions, display properties and cultural background of each participant.

Most observers conducted the experiment in typical domestic (28%) lighting conditions followed by dark (24%), mid morning/afternoon daylight (18%), typical office (17%), north sky daylight (7%) and noon daylight (6%) conditions. The white point of the monitor was described in most cases as neutral white (50%), as bluish white (29%), as warm white (15%) and as yellowish white (6%). The area outside the monitor filling the visual field of the observers, known as surround, was described as bright (10%), average (45%), dim (19%) and dark (26%).

A more detailed description of the procedure and verification of the results against previous studies conducted in controlled viewing conditions[24,25] can be found in Mylonas and MacDonald.[19]

Results and Analysis

For this study, we analyzed data in English from observers over 16 years of age with normal trichromatic colour vision (83%). Original responses were filtered for spelling mistakes, hyphenations and comma separators; words in parenthesis were treated as multiword colour expressions while various typographic conventions were removed. Excluded were incomplete responses or responses in languages other than English. The refined dataset resulted in 5428 observations, including 1166 unique colour words. Of these, 53% involved single-word (monolexemic) responses, 42% two-word responses, 4% three-word responses and 1% four or more words. The 11 BCTs occurred in 29% while non-BCTs occurred in 24% of cases (Figure 1).

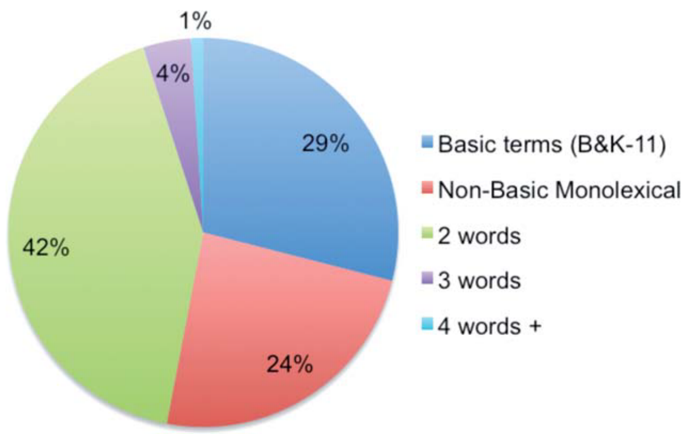

Figure 1. Number of words in colour names collected in online experiment.

We analyzed the data in terms of frequency of the term usage, consensus among genders, consistency of responses, response times, denotative volume in colour space and a validation metric against a parallel online experiment.[28] The ability of each measure to separate BCTs from non-BCTs was quantified by an one-tail Wilcoxon rank sum test at the 5% significance level ($p>0.05$) using the five lowest-ranked B&K terms and the five top-ranked non-basic terms in each test. We expected that BCTs would tend to rank higher than nonbasic terms and hence would be distinguishable.

Frequency

The most frequent 30 monolexemic colour terms found in our study are shown in Figure 2. The colour term with the highest frequency was purple followed by pink, blue, green, yellow and brown. Non-BCTs turquoise, lilac and violet occurred in the 7th, 8th and 9th positions. The least frequent basic terms were red (13th) and white (19th). Magenta, mauve, cyan and fuchsia were found in the 14th, 15th, 16th and 17th positions. No significant advantage in frequency was found for the last five BCTs of B&K over the first five non-basic terms ($p=0.85$). Responses to repeated colour samples were excluded from this measure (see consistency metric).

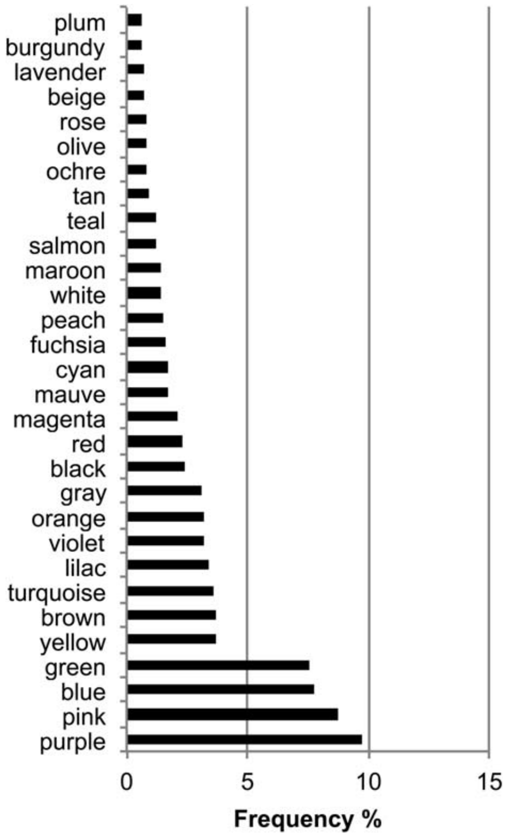

Figure 2. Frequency of occurrence of monolexemic colour terms.

Consensus Among Genders

In a previous study,[35] we explored gender differences in colour naming and showed that females demonstrated more elaborated colour vocabulary and faster responses than males. In this study, we measured the consensus between genders to quantify the agreement for the name of a colour across groups of speakers of a language. Figure 3 shows the results expressed as colour differences (ΔE^*_{ab}) between the centroids, which specify the center of the region of samples named by each monolexemic colour term by females and males.

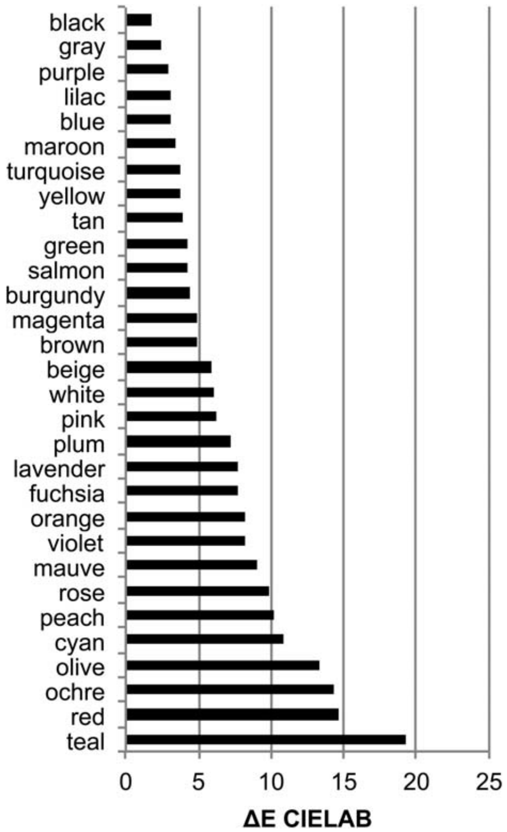

Figure 3. Gender differences between the centroids of monolexemic colour terms measured by Euclidian distances in CIELAB (ΔE^*_{ab}).

Black, gray and purple were the colour names with the best agreement between males and females. No advantage of the BCTs was found over non-BCTs ($p=0.99$), as lilac, maroon, turquoise and tan were in the first 10 positions. Among the BCTs, red and orange showed the worst agreement between genders.

Response Time

The RTs were calculated with cut-off of 3 standard deviations from the mean time required to name each colour sample. Figure 4 shows the response time for each of the most frequent colour terms in seconds. Red was found to be the term with the fastest response followed by blue, white and green. Teal, peach and olive were non-BCTs with the fastest responses. BCTs with the longest RTs were purple and gray.

Non-BCTs with the longest RTs were lavender, turquoise and plum. The response time metric produced significant differences between B&K's BCTs and non-BCTs (p=0.048) and we replicated previous findings.[24–26]

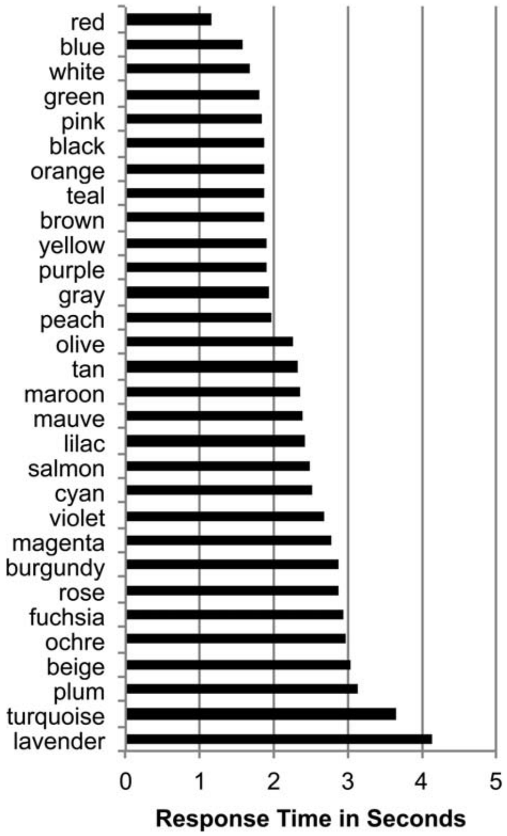

Figure 4. Mean response times (seconds) of the most frequent monolexemic colour terms.

Consistency of Responses

In the online colour naming experiment, one randomly selected colour sample was repeated twice in each session to measure the consistency of unconstrained colour naming responses. In other words, consistency measures the agreement for the name of a colour sample presented twice by a single observer. Following the consistency measures of Guest and Laar,[36] the overall consistency for the entire colour names was 36% while by comparing only their hue component and excluding the modifiers was 67%. Participants were not informed about the repeated colour sample and each repetition was separated by more than 10 colour identifications. Monolexemic colour terms with the highest consistency are shown in Figure 5.

The most consistent BCTs were blue, green and purple. Lilac, ochre, teal and turquoise were the most consistent non-BCTs. The least consistent BCTs were orange and white. In the test data, observers did not repeat 11 of the 30 most frequent monolexemic colour terms. No significant differences in consistency were found between BCTs and non-BCTs (p=0.68).

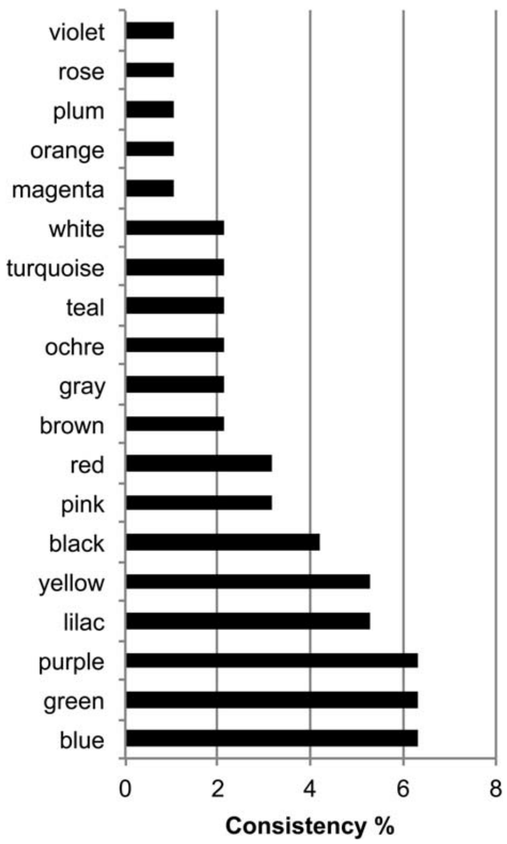

Figure 5. Consistency of monolexemic colour terms.

Denotative Volume

To establish the volume of each colour category, we used a parametric model based on Maximum a Posteriori (MAP)[37] and a mixture of Gaussian distributions of the most frequent English colour terms as a learning set (n=30), to label all 1729 samples of the Munsell Renotation Dataset located within the sRGB gamut. We used a 'round trip' set of transformations to determine which colour samples were located outside the sRGB gamut.[38]

For each colour term y from a set of the most frequent monolexemic colour terms $y_1,....,y_T$ responded by the participants of our experiment, including the repeated responses, we calculated the empirical mean μ_y and variance-covariance matrix Σ_y of test colour patches $x_1,....,x_n$. The probability density function could be then estimated by:

$$\hat{f}_{norm}(x \mid y) = const_y \exp\left(-\frac{1}{2}\left(x - \mu_y\right)^{\mathrm{T}} \Sigma_y^{-1}\left(x - \mu_y\right)\right), \; x \in \{x_1,...,x_n\} \qquad [1]$$

where: x is a test colour specified by the triplet $x = (x_{(L)}, x_{(a)}, x_{(b)})^T$ and $const_y$ is a normalizing factor depending on μ_y, Σ_y and $x_1,....,x_n$ which ensures that the sum of the whole probability distribution is equal to 1. It is noted that the exponent is equal to minus half the squared Mahalanobis distance between x and μ_y. Using the Bayes' theorem, the MAP estimator is defined as:

$$\hat{y}_{MAP}(x) = \underset{y \in \{y_1,\ldots,y_T\}}{\arg\max} \left(\frac{\hat{f}_{norm}(x \mid y) \cdot \hat{f}(y)}{P(X = x)} \right) \qquad [2]$$

MAP favours colour names with high probability $\hat{f}(y)$ or high normalization factor $const_y$ to maintain congruence between observed and predicted data. This means that μ_y is not necessarily equal to the mean of $\hat{f}_{norm}(x \mid y)$ and frequent and consistent colour categories tend to subsume less common and inconsistent neighbour categories. Figure 6 shows (A) the classified Munsell sampling in the CIELAB colour space and (B) the volume of each identified colour name. The volume was calculated as the percentage of colour samples (out of 1729) named by the colour term in question.

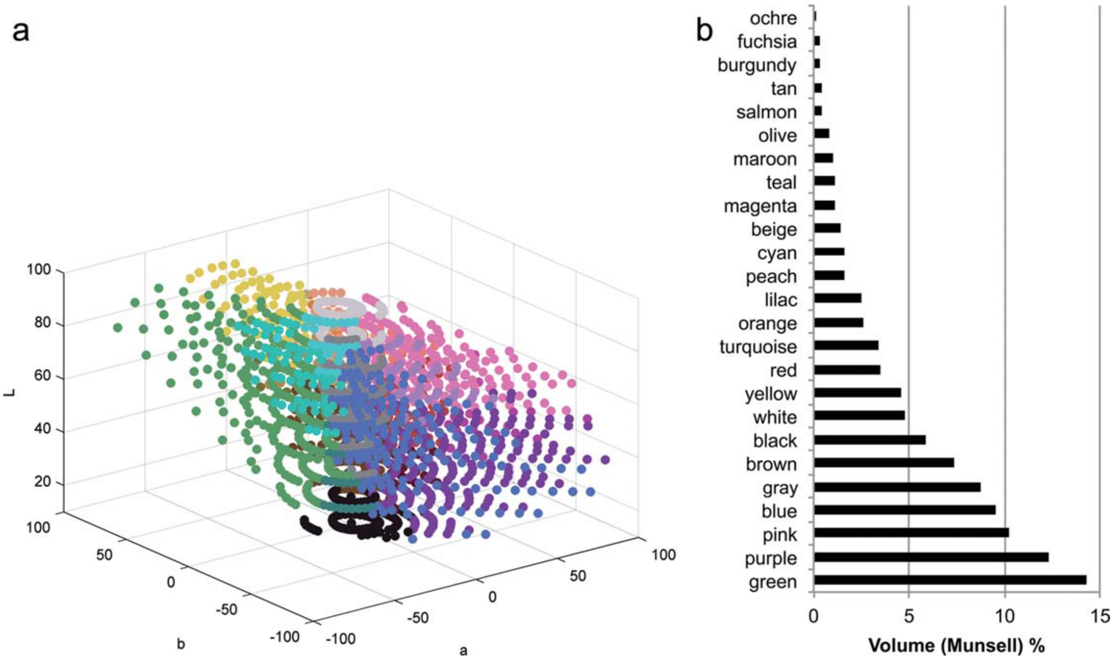

Figure 6. Volume of 25 most dominant monolexemic colour terms in the classified Munsell sampling (A: sRGB gamut in CIELAB, 1729 samples & B: % number of samples).

The MAP algorithm identified 25 predominant colour categories as larger categories subsumed *violet, lavender*, *rose, plum* and *mauve*. The largest category was *green* followed by nine BCTs. *Turquoise* was found in the 11th position followed by *orange*, *lilac* and *peach*. The difference in this measure was significant (p=0.008) and basic terms appear to cover larger volumes of the Munsell system than non-basic terms. To establish whether our volumetric results were influenced by the particular sampling of the Munsell Renotation Dataset, we used the same MAP algorithm and same learning set (n=30) to classify 399 samples of the radial sampling of the OSA space located within the sRGB gamut [39]. Figure 7 shows (A) the classified radial OSA sampling and (B) the predicted volume of each identified colour name.

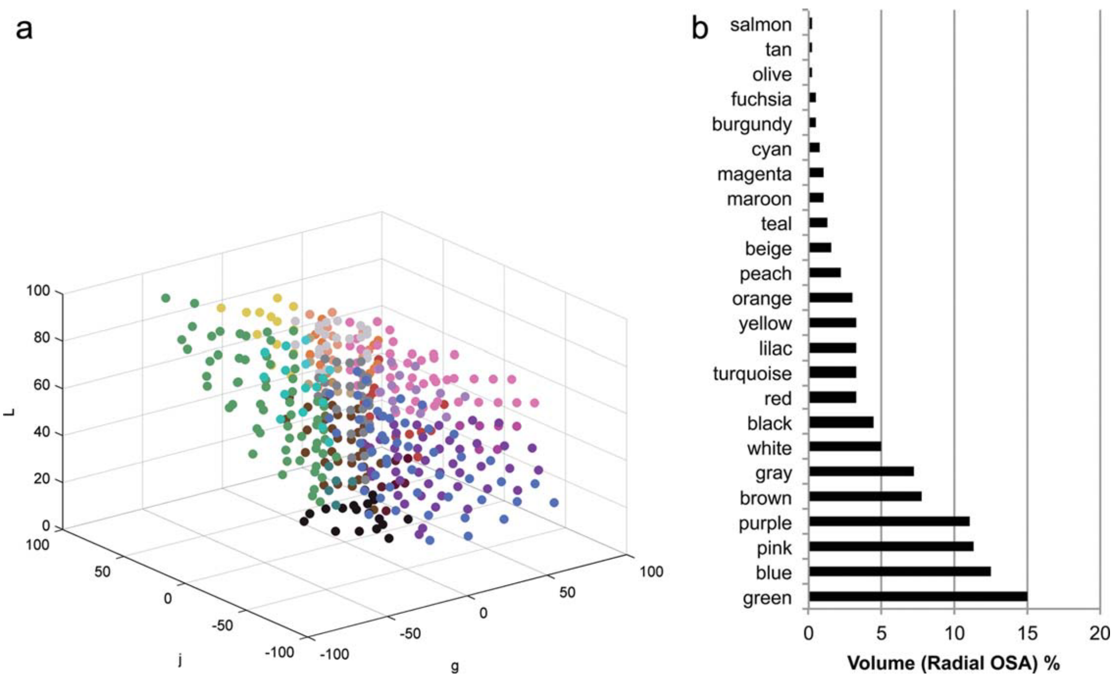

Figure 7. Volume of most dominant monolexemic colour terms in the radial OSA sampling (A: sRGB gamut, 399 samples in CIELAB & B: % number of samples).

In this case the metric identified 24 predominant colour categories in the Radial OSA sampling as larger categories subsumed *lavender, mauve, ochre, plum, rose* and *violet*. The category with the largest volume was *green*, followed by eight BCTs. *Turquoise* and *lilac* were found in the 10th and 11th positions respectively followed by *yellow, orange* and *peach*. Notably, for both tested samplings of colour space, the 13 identified colour terms with the largest volume were the 11 BCTs plus *turquoise* (11th or 10th positions respectively) and *lilac* (13th or 11th positions respectively). Although using a different colour order system has influenced the results, the metric based on volume produced again significant differences between basic and non-basic colour terms (p=0.04) for the Radial OSA sampling.

Inter-experimental agreement

To validate our results against other studies, we measured the inter-experimental agreement between the outcomes of our study and Moroney's online experiment [28]. Figure 8 shows the colour differences (ΔE^*_{ab}) between centroids of the 30 most frequent colour terms.

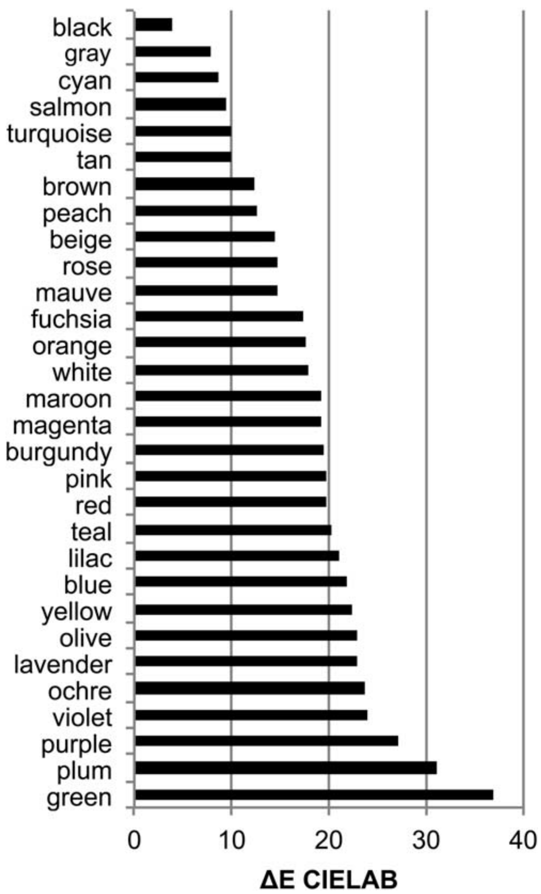

Figure 8. Inter-experimental agreement between the present study and Moroney's (2003) online experiment, measured by the colour differences (ΔE^*_{ab}) between coordinates of centroids for the most frequent monolexemic colour terms.

Black was found to be the colour term with the smallest colour difference, followed by *gray* and *cyan*. No advantage was found for the BCTs over non-basic colour terms (p=0.99). A complement to the 3D colour difference (ΔE^*_{ab}) between centroids is the difference in centroid hue angle (Δh_{ab}), which is a measure less dependent on the lightness differences caused by variations in display luminance and/or room lighting. Figure 9 shows agreement of the two online experiments in terms of hue angle (achromatic BCTs were excluded).

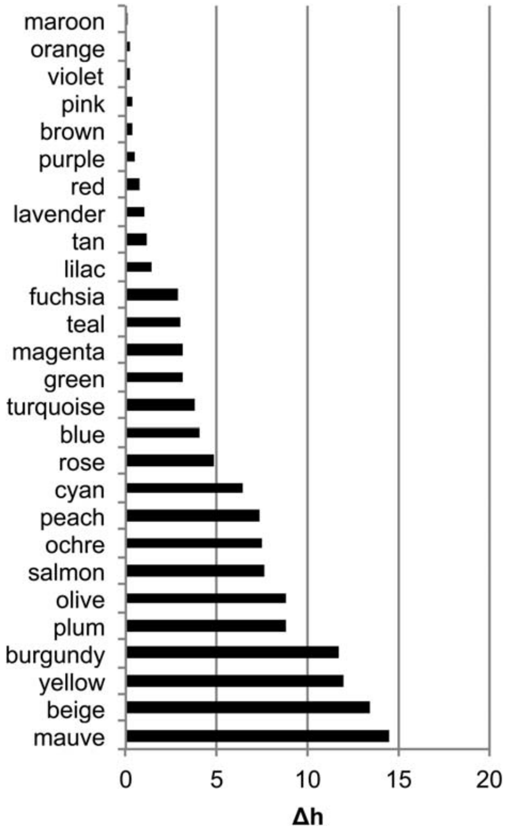

Figure 9. Inter-experimental agreement of centroid hue angle (Δh_{ab}) for the most frequent achromatic monolexemic colour terms.

Maroon was the colour term with the smallest hue angle difference, followed by *orange*, *violet*, *pink*, *brown, purple,* and *red*. For all of these terms the hue angle difference, $\Delta h_{ab} < 1$, was very small, indicating a close agreement between the results of the two online experiments. The BCT with the largest colour difference was *yellow*. *Lilac* and *turquoise* were found in the 10^{th} and 15^{th} positions respectively. No significant advantage of the BCTs was found over the non-basic colour terms ($p=0.92$).

Composite index of basicness

Behavioural measures for separating basic from non-basic terms varied in their effectiveness. To encapsulate all of our measures in a single metric, we calculated the means of the ranks for each colour term [12] across all six appropriate measures. We excluded the hue angle difference metric because it does not address achromatic colour terms. We also excluded the second verification metric of volume using the radial OSA sampling, to avoid including in our calculations the same measure twice. To address the issue of ordinality for practical purposes, we first replaced each colour name with its rank and then transformed the range of each variable onto the unit interval [0,1] by dividing each variable by the number of the corresponding ordered categories after subtracting the minimum value [40]. The combined index of basicness for each colour term is shown in Figure 10 by order of the mean rank. Low values indicate a high degree of basicness, where the colour term was near the top of the ranking list in the majority of measures.

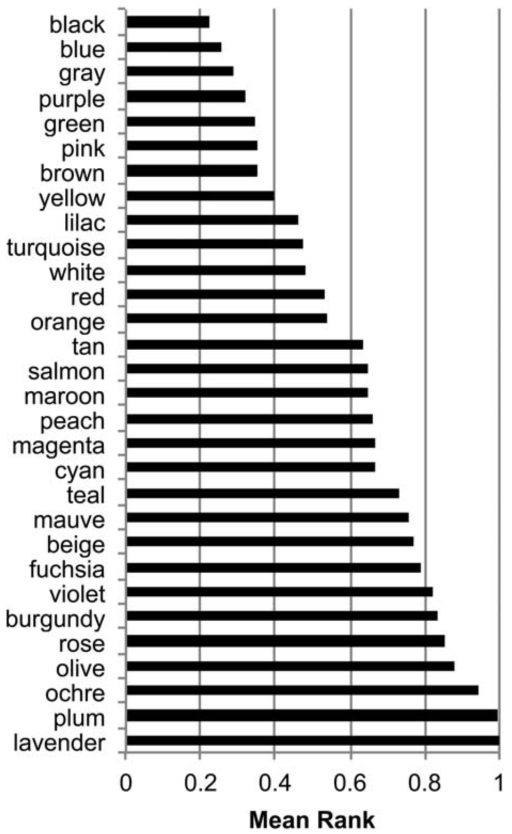

Figure 10. Index of basicness for the most frequent monolexemic colour terms.

The 11 BCTs of B&K were found in the top positions but the non-basic *lilac* and *turquoise* were also included in the first top ten terms. *Black* was ranked at the top of the list followed by *blue* and *gray*. *White, red* and *orange* occurred in the 11th, 12th and 13th positions respectively. The metric produced no significant differences between the last five basic colour terms of B&K and the first five non-basic terms (p=0.09).

To determine the threshold index value between basic and non-basic colour terms we constructed two agglomerative clusters from the index of basicness of each monolexemic term. Basic colour terms sharing a similar index of basicness should be grouped in the first cluster while non-basic colour terms should be grouped in the second cluster. We tested four different distance metrics: Euclidean, city block, Minkowski and Mahalanobis. All four metrics identified 13 basic colour terms in the first cluster, including *lilac* and *turquoise* with the 11 basic colour terms of B&K. These were the same 13 colour terms ranked highest by the volume metric for both colour order systems.

To verify statistically the identified threshold index we assessed whether the lowest five ranked basic terms (8th-13th) are significantly separated from the first five non-basic terms (14th-19th) in the mean rank. Including *lilac* and *turquoise* in basic terms produced significant differences (p=0.004). The same results were obtained when we tested the volume metrics for 13 basic terms for the Munsell ($p=0.004$) and OSA colour order systems ($p=0.004$).

Augmenting the precision of colour naming in colour spaces

Having additional basic colour terms can improve the precision of colour naming algorithms, as shown in Figure 11. The coordinates of the centroids were used to colour each colour category. The performance of the probabilistic model based on MAP [37] with training sets consisting of thirteen basic colours terms to segment a synthetic image [41] was superior to the performance with eleven basic terms.

Turquoise covers a large region between *green* and *blue*, while *lilac* is concentrated in the pale and high lightness area of *purple*. *Red*, similar to *purple*, covers only the high-saturated regions of the synthetic image. Note that neither *black* nor *white* is present in the synthetic image because it is constructed as a diagonal slice through colour space.

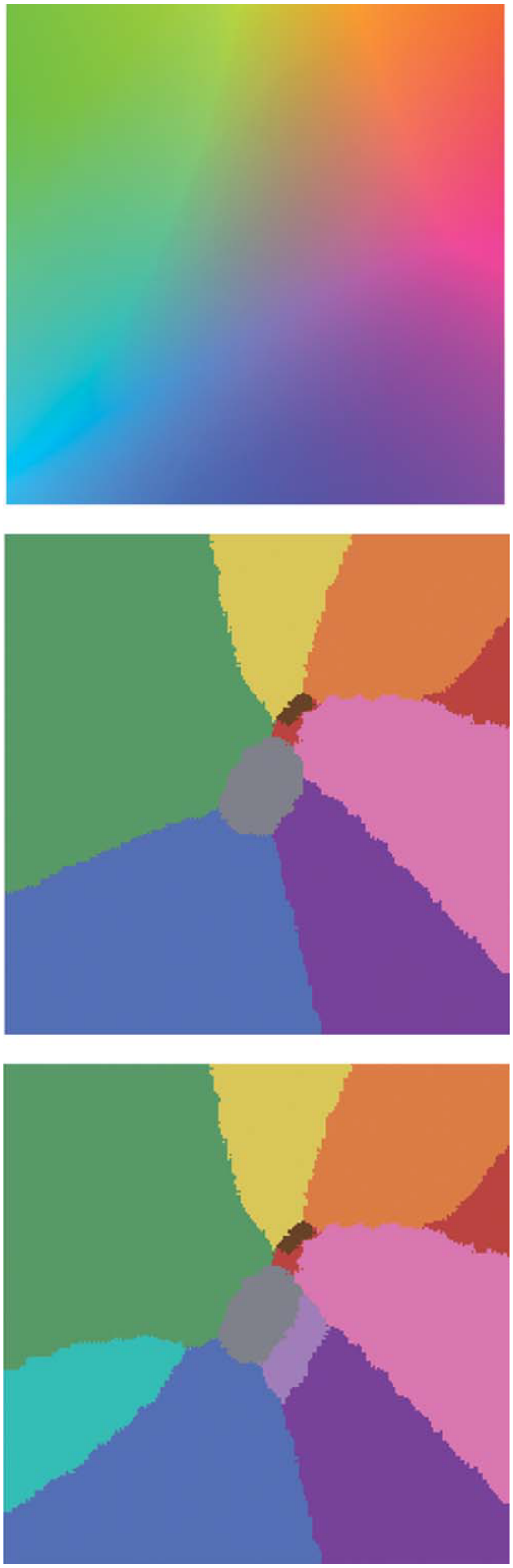

Figure 11. Segmentation of a synthetic image in CIELAB: (top) Original image; (middle) MAP-training set of 11 basic terms; (bottom) MAP-training set of 13 basic terms.

Discussion and Conclusions

This study has examined whether the number of basic colour terms could be extended by analysis of a large number of responses from an online colour naming experiment with a set of behavioural measures. Our long-standing aim has been to promote colour communication within and across cultures and to improve the precision of colour naming in perceptual colour spaces. A larger colour vocabulary can improve the accuracy with which a colour name communicates the referent colour and can make colour communication easier [42] over time. Our findings suggest the extension of the 11 basic colour terms in English to 13 including the terms *lilac* and *turquoise*.

By translating the notion of basicness into a set of criteria, we have used the performance of a set of tests to separate basic from non-basic colour terms. In terms of frequency of occurrence, *turquoise* ranked 7th while *lilac* ranked 8th (Figure 2). The consensus among genders metric revealed (Figure 3) that *lilac* was the 4th and *turquoise* was the 7th colour term. Regarding speed of responses, however, *lilac* was ranked 18th and *turquoise* 29th (Figure 4). A possible explanation for the notably long response time for *turquoise* is that participants had difficulty in spelling the word. In total 32 different spellings were found in the raw responses before any corrections or colour vision test filters were applied. *Lilac* was the 4th most consistent colour term in our data while *turquoise* was 11th. In the Munsell Renotation Dataset situated in the sRGB gamut, *lilac* and *turquoise* covered the 13th and 11th largest volumes respectively. This was similar to the radial sampling of the OSA space, where *lilac* and *turquoise* were found in the 10th position (Figures 6-7). The validity of our results was assessed as the agreement between two online experiments. *Lilac* was positioned 21th and *turquoise* 5th using Euclidean distances in CIELAB (ΔE_{ab}), while for hue angle differences (Δh_{ab}) *lilac* was found in the 10th and *turquoise* in the 15th positions (Figures 8-9).

The six separate metrics were combined to provide a composite ‘index of basicness’ as a mean rank [12]. Next to the classical 11 English BCTs, *lilac* and *turquoise* occupied the 9th and 10th index positions of most frequent monolexemic colour terms (Figure 10). In terms of index differences, separation from the lowest ranked basic colour terms, *white, red* and *orange*, was moderate but there was a considerable jump in index value to the following non-basic term *tan.* This was verified by separating basic from non-basic terms into two agglomerative clusters using their index of basicness and four different metrics of distance, all producing the same result – the same 13 colour terms that were ranked first in the volume metric for both colour order systems.

Most metrics did not produce significant results on their own, except the measures of response time and denotative volume using a classifier of *Maximum a Posteriori* (MAP). MAP employs a mixture of Gaussian distributions of colour names and the frequency of their occurrence as a prior distribution to maintain a balance between observed and predicted data. In other words, this synthetic observer combines the distribution of unique colour samples identified with the same colour name with the frequency and consistency of this name. Many colour names are subsumed by larger colour categories. For example, *mauve* is a widely used term in English but *purple* is used even more frequently and its *denotata* are larger in volume. As a result the region associated with *mauve* identified as *purple*, whereas it is separated from *pink* and *lilac* (see Figure 6 and Figure 7). Volume measures are dependent on the geometry of the sampling grids, but we have tested two different samplings of the colour space and found that both produced significant separation of the 11 basic colour terms of B&K. The differences between the performances of the metric on the two different grids can be explained by the cylindrical structure of the Munsell system that over-samples lower-saturation regions of colour space. To represent perceptual uniform hue spacing, the Munsell collection also includes a larger number of purple hues. The performance of the metric was improved for both grids (p=0.004) when we tested it with 13 basic terms; including *lilac* and *turquoise*.

Sturges and Whitfield [25] suggested *cream, turquoise* and *lilac* as the highest-ranking non-basic colour terms but the difference of *lilac* from the BCTs was significantly greater than in our study. *Cream* was found to be the 34th most frequent colour term in our data so further investigation is needed to test its status. *Tan* was found in the 14th position of our index of basicness but it was not qualified as a BCT. Davies *et al.*[26] reported *turquoise, mauve* and *lilac* as the most frequent non-basic terms. Jrassaiti *et al.*[43] reported 14 consensual colour terms, including *peach, lilac* and *turquoise* in addition to the 11 BCTs.

Zimmer [44] suggested *turquoise* as an additional universal BCT in his continuous model for 'basicality', at least for German speakers. Witzel and Gegenfurtner [45] suggested the existence of a *turquoise* category between *blue* and *green* for both German and non-German observers that is not equivalent to any of the BCTs of B&K. Walter [46] reported that *turquoise* categorization is consistent even when viewed through strong yellow-orange filters and its range is expanded towards the *blue* and *green* categories.

Paramei [47] and Androulaki *et al.*[48] found that Russian and Greek languages both have twelve BCTs, differentiating 'light blue' from 'dark blue'. We were therefore expecting to see the division of *blue* in the observer responses into two basic terms, along the lines proposed by Jameson [49]. In a previous study [19], where we did not constrain our analysis to monolexemic terms, we found English *light blue* and *sky blue* in the 11th and 20th positions respectively. *Cyan* was the most frequent monolexemic colour term used by our observers to describe this region of colour space but it was predominantly used by males (perhaps indicating their knowledge of colour names for subtractive printing primaries) and less often by females.

Goodman [50] coined the word *grue* to express a philosophical conundrum relating to induction, namely "What colour would the metonymic *emerald* be if one day someone were to discover a blue emerald?" The term *grue* was subsequently adopted by Kay [51] to represent the combination of green and blue hues, when analysing the evolutionary sequence in which one or other colour term first appeared in a language. He described this as a composite colour category, the fuzzy union of two primary colour categories. Kay and McDaniel [22] developed the theory to predict that the foci of *grue* should be bimodal, with some languages showing a focus corresponding to *green* and others to *blue*. But in analysis of the World Color Survey (WCS) data, Lindsey and Brown [52] found that two patterns of *grue* naming could be discerned in the 110 languages of the WCS. One group placed focal *grue* near a single primary category focus, either green or blue, whereas the second group seemed to combine green and blue into a single perceptual category and placed focal *grue* near the centre of the region. We contend that the dichotomy arose because *green*, *grue* and *blue* are actually three separate colour categories, each with its own perceptual identity. The recognition of *grue* as a basic colour term, under the synonym *turquoise*, provides a satisfactory explanation of both colour naming behaviour and the evolution of colour language.

In a recent paper Lindsey and Brown [53] reported the current state of evolution of the colour lexicon of American English, and identified four candidates of non-basic colour terms to join BCTs: *teal, peach, lavender,* and *maroon*. We note the close similarity of *teal* to *turquoise* and *lavender* to *lilac*. Regarding the origin of new basic colour categories, the authors suggested that both emergence[11] and successive differentiation[54] takes place when new basic colour terms arise. We support this view, as *lilac* appears to partition the large colour category of *purple* while *turquoise* appears at the boundary between *green* and *blue*. *Peach* and *maroon* were found in the 17th and 15th positions in our index of basicness but were not qualified as BCTs.

In conclusion, we do not propose that thirteen is the definitive number of basic colour terms in English. Rather we observe that colour language is in a constant state of flux and continues to unfold new basic, or commonly shared, colour terms as the cultural need to communicate about colour evolves. Future plans include the extension of the research into other languages and examination of the role of primary and non-primary colour terms.

Acknowledgements
We thank the thousands of participants in our online colour naming experiment and all the volunteers who have contributed in this project. We also thank Prof. Galina Paramei and Prof. Jules Davidoff for valuable comments on earlier drafts of the manuscript and Dr Michael Studer for his help towards the statistical analysis.

References

1. Pointer MR, Attridge GG. The number of discernible colours. Color Res Appl 1998;23:52–54.

2. Mojsilovic A. A computational model for color naming and describing color composition of images. IEEE Trans Image Process 2005;14:690–699.

3. Moroney N, Obrador P, Beretta G. Lexical image Processing. In Proceedings of IS&T/SID 16th Color Imaging Conference, Springfield VA: IS&T; 2008. p 268–273, San Antonio, TX.

4. Heer J, Stone M. Color naming models for color selection, image editing and palette design. In Proceedings of ACM, 2012. p 1007–1016, Austin, TX.

5. Lammens JMG. A Computational Model of Color Perception and Color Naming. PhD Thesis. State University of New York at Buffalo; 1994

6. Benavente R, Vanrell M, Baldrich R. Parametric fuzzy sets for automatic color naming. J Opt Soc Am A 2008;25:2582–2593.

7. Motomura H. Categorical color mapping for gamut Mapping. In Proceedings of 5th IS&T/SID Colour Imaging Conference, Springfield VA: IS&T; 1997. p 50–55, Scottsdale, Arizona.

8. Lin H, Luo MR, MacDonald LW, Tarrant AWS. A cross-cultural colour- naming study. Part III—A colour-naming model. Color Res Appl 2001;26:270–277.

9. Davidoff J. Cognition Through Color. Cambridge, MA: MIT Press; 1991.

10. Gegenfurtner K, Rieger J. Sensory and cognitive contributions of color to the recognition of natural scenes. Curr Biol 2000;10:805–808.

11. Berlin B, Kay P. Basic Color Terms: Their Universality and Evolution. Berkeley, CA: Cambridge University Press; 1969/1999.

12. Corbett GG, Davies IRL. Establishing basic colour terms: Measures and techniques. In: Hardin CL, Maffi L, editors. Color Categories in Thought and Language. Cambridge UK: Cambridge University Press; 1997. p 197–223.

13. Kay P, Regier T. Resolving the question of color naming universals. Proc Natl Acad Sci USA 2003;100:9085–9089.

14. Bimler D. Are color categories innate or internalized? Hypotheses and implications. J Cognit Cult 2005;5:265–292.

15. Griffin LD. Optimality of the basic colour categories for classification. J Roy Soc Interface 2006;3:71–85.

16. Regier T, Kay P, Khetarpal N. Color naming reflects optimal partitions of color space. Proc Natl Acad Sci USA 2007;104:1436–1441.

17. Dummett M. Frege: Philosophy of Language, Cambridge, MA: Harvard University Press; 1973.

18. MacDonald LW, Mylonas D. Edible colour Names. In Proceedings of AIC, 2010. p 196–199, Mar del Plata, Argentina.

19. Mylonas D, MacDonald LW. Online colour naming experiment using munsell colour Samples. Proceedings of 4th IS&T/SID on Colour in Graphics, Imaging, and Vision, 2010. p 27–32, Joensuu, Finland.

20. Mylonas D, MacDonald L. Colour naming for colour communication. In: Best J, editor. Colour Design: Theories and Applications. Cambridge, UK: Woodhead Publishing; 2012. p 254–270

21. Brown RW, Lenneberg EH. A study in language and cognition. J Abnorm Soc Psychol 1954;49:454–462.

22. Kay P, McDaniel CK. The linguistic significance of the meanings of basic color terms. Language 1978;54:610–646.

23. Hering E. Outlines of a Theory of Light Sense. Translated from German by Leo Hurvich and Dorothea Jameson 1964. Cambridge: Harvard University Press; 1892.

24. Boynton RM, Olson CX. Locating basic colors in the OSA space. Color Res Appl 1987;12:94–105.

25. Sturges J, Whitfield A. Locating basic colours in the munsell space. Color Res Appl 1995;20:364–376.

26. Davies IRL, Corbett GG, Laws G, McGurk H, Moss AE, Smith MW. Linguistic basicness and colour information processing. Int J Psychol 1991;26:311–327.

27. Kerttula S. Relative basicness of color terms: Modeling and measurement. In: MacLaury RE, Paramei GV, Dedrick D, editors. Anthropology of Color: Interdisciplinary Multilevel Modeling. Amsterdam/Philadelphia: John Benjamins; 2007. p 151–170.

28. Moroney N. Unconstrained web-based color naming experiment. In Proceedings of SPIE, Vol. 5008, 2003. p 36–46, Santa Clara, CA.

29. Davidoff J, Davies I, Roberson D. Colour categories in a stone-age tribe. Nature 1999;398:203–204.

30. Roberson D, Davidoff J, Davies IRL, Shapiro LR. Color categories: Evidence for the cultural relativity hypothesis. Cogn Psychol 2005;50: 378–411.

31. Drivonikou GV, Kay P, Regier T, Ivry RB, Gilbert AL, Franklin A, Davies IRL. Further evidence that whorfian effects are stronger in the right visual field than the left. Proc Natl Acad Sci USA 2007;104:1097–1102.

32. Gage J. Color and Culture: Practice and Meaning from Antiquity to Abstraction. Boston: Little, Brown and Company; 1993.

33. Hardin CL, Maffi L. Introduction. In: Hardin CL, Maffi L, editors. Color Categories in Thought and Language. Cambridge UK: Cambridge University Press; 1997. p 1–18.

34. Barbur JL, Harlow AJ, Plant GT. Insights into the different exploits of colour in the visual Cortex. Proc R Soc London B Biol Sci 1994;258: 327–334.

35. Mylonas D, Paramei GV, MacDonald LW. Gender Differences in Colour Naming. In: Anderson W, Biggam CP, Hough C, Kay C, editors. Broad Spectrum of Colour Studies. Amsterdam/Philadelphia: John Benjamins;2014. p 225–239.

36. Guest S, Laar DV. The structure of colour naming space. Vision Res 2000;40:723–734.

37. Mylonas D, MacDonald LW, Wuerger S. Towards an online color naming model. In Proceedings of 18th IS&T/SID Colour Imaging Conference, Springfield VA: IS&T; 2010. p 140–144, San Antonio, TX.

38. Kang HR. Computational Color Technology. Bellingham, Washington: SPIE Press; 2006.

39. Moroney, N. A radial sampling of the OSA uniform color scales. In Proceedings of 11th IS&T/SID Colour Imaging Conference, Springfield VA: IS&T; 2003, Scottsdale, Arizona.

40. Kaufman L, Rousseeuw PJ. Finding Groups in Data: An Introduction to Cluster Analysis. Hoboken, N.J: Wiley-Interscience; 2005.

41. Weijer J, Schmid C, Verbeek J. Learning color names from Real-world Images. IEEE Trans Image Process 2009;18, Minneapolis, MN; p. 1–8.

42. Lantz D, Stefflre V. Language and cognition revisited. J Abnorm Soc Psychol 1964;69:472–481.

43. Jraissati Y, Wakui E, Decock L, Douven I. Constraints on colour category formation. Int Stud Philos Sci 2012;26:171–196.

44. Zimmer AC. What really is turquoise? a note on the evolution of color terms. Psychol Res 1982;44:213–230.

45. Witzel C, Gegenfurtner KR. Is there a lateralized category effect for color? J Vis 2011;11:16.

46. Walter S. Perceiving “grue”: Filter simulations of aged lenses. In: Biggam CP, Hough CA, Kay CJ, Simmons DR, editors. New Directions in Colour Studies. Amsterdam: John Benjamins; 2011. p 329–342.

47. Paramei GV. Singing the Russian blues: An argument for culturally basic color terms. Cross Cult Res 2005;39:10–38.

48. Androulaki A, G^omez-Pesta~n N, Mitsakis C, Davies I. Basic colour terms in modern Greek: Twelve terms including two blues. J Greek Ling 2006;7:3–47.

49. Jameson KA. Why GRUE? an interpoint-distance model analysis of composite color categories. Cross Cult Res 2005;39:159–204.

50. Goodman N. Fact, Fiction and Forecast. Cambridge, MA: Harvard University Press. Chapter III. Cited by Barker SF, Achinstein P (1960). On the new riddle of induction. The Philosophical Review; 1955. p 511–522.

51. Kay P. Synchronic variability and diachronic change in basic color terms. Lang Soc 1975;4:257–270.

52. Lindsey DT, Brown AM. Sunlight and “blue”: The prevalence of poor lexical color discriminations within the “grue” range. Psychol Sci 2004; 15:291–294.

53. Lindsey DT, Brown AM. The color lexicon of American English. J Vis 2014;14:1–25.

54. Levinson SC. Yélî Dnye and the theory of basic color terms. J Ling Anthropol 2000;10:3–55.