

# The Biological Basis of the Experience and Categorization of Colour

*Semir Zeki, Alexandre Javier and Dimitris Mylonas*

Laboratory of Neurobiology,

University College London

[Corresponding author: S. Zeki. S.zeki@ucl.ac.uk]

## Abstract

We used the Land Colour Mondrian experiments in a Bayesian context to test the degree to which subjects vary in categorizing the colour of different patches, when each patch is made to reflect light of the identical wavelength-energy composition. The brain uses a ratio-taking mechanism to determine the ratio of light of every waveband reflected from a surface and from its surrounds. Our (Bayesian) hypothesis was that this ratio-taking mechanism is similar in all humans and therefore leads to a constant categorization of colours that differs little between them. The similarly categorized colours are the initial priors, with initial hues attached to them. Twenty subjects of different ethnic and cultural backgrounds, for all but one of whom English was not the primary language, viewed 8 patches of different colour in two Mondrian displays; each patch, when viewed, was made to reflect identical ratios of long-, middle- and short- wave light. Subjects were asked to match the colour of the viewed patch with that of the Munsell chip coming closest in colour to that of the viewed patch, without using language. In terms of hue, there was less variability in matching warm hues than cool ones. In terms of colour categorization, there was little variability overall. We take the lack of significant variability between subjects in the matches made as a pointer to similar computational mechanisms being employed in different subjects to perceive colours, thus permitting them to assume that their categorization of colours has universal agreement and assent.

## Keywords

colour vision, Bayesian operations, constant categorization

## Data Accessibility

Data underlying this article can be accessed on Open Science Framework at [osf.io/ge3uq](https://osf.io/ge3uq), and used under the CC-BY Attribution 4.0 International.

## Introduction

It is trite neurobiology to say that one of the primordial functions of the brain is to acquire knowledge. Yet this raises a fundamental issue of huge importance, namely the extent to which the knowledge acquired through brain mechanisms by one individual is identical to that acquired by another or others, thus allowing the acquiring individual to assume reasonably that there is universal assent to the knowledge and experience acquired by him or her. The question resolves, therefore, around asking what conditions, if any, enable all individuals, irrespective of their ethnic, cultural or educational status, to share the same experience and knowledge under the same conditions.

There has been much philosophical debate about this subject, which we do not delve into here in any detail. Rather, accepting Immanuel Kant's (1781) statement that "perceptions without concepts are blind", we work on the assumption that all sensory inputs are interfaced through, **or based on**, brain concepts, of which, we believe, there are two kinds: acquired (synthetic) concepts and inherited one (Zeki, 2009). The former, among which we include concepts related to human-made objects such as utensils, cars and furniture, lead to experiences and knowledge that can differ profoundly between individuals even when experienced under identical conditions, and that can be modified throughout postnatal life. This makes it unsafe for the experiencing individual to assume that his or her experience will be identical or even similar to that of others. The latter, inherited, concepts lead to knowledge which, to a greater or lesser extent, is similar in all individuals and which is resistant to change through experience, thus making it easy for one individual to assume that others would share the same or nearly identical experience under similar conditions. In the

work reported here we concentrate on inherited concepts alone and on one of the most extreme examples of it, namely colour vision, which we approach in a Bayesian context.

We could of course have used terms like algorithm or program, which have found wide usage in computational neurobiology, rather than the term ‘concept’ to describe the brain’s inherited ratio-taking mechanisms for generating colours. We prefer to use the term ‘concept’ here for two reasons; partly because it perpetuates the term used initially by Immanuel Kant when he wrote that all experiences, except time and space, must be interfaced through concepts, and partly because, in the Bayesian context in which we write, there are other inherited ‘concepts’ which have cross-cultural validity, such as that of ‘unity-in-love’, with which the term algorithm does not sit so easily and for which the term concept seems better suited (Zeki, 2009; Zeki & Chen 2019 for a discussion of the Bayesian brain).

The Bayesian approach supposes that initial *priors*, or initial beliefs, generate *posteriors* through experience, experimentation and learning. The posteriors thus generated (through experience and learning) have (modified) beliefs attached to them and therefore modify inference and behaviour; the (modified) posteriors can then act as new priors from which, through experience and experimentation, new posteriors are generated, and so on, iteratively (Dayan *et al.*, 1995; Rao & Ballard, 1999; López *et al.*, 2014; Friston *et al.*, 2011; Pouget *et al.*, 2013) (*inter alia*). Our hypothesis in this work is that, in colour vision, the initial *prior* generated from the brain’s inherited ratio-taking mechanisms (Land, 1986) leads to (a) constant colour categorization (Zeki *et al.*, 2017), to which (b) a given initial hue is attached; the initial prior generated from the brain’s ratio-taking mechanism has, therefore, two constituents; we refer to these two constituents collectively as the initial biological or. *B prior*. However, unlike Bayesian priors in general, one of the two constituents of this initial biological prior, the constant colour categorization, is substantially the same in all individuals and is not easily modified even with extensive experience and learning. Hence, an individual can legitimately assume that the colour categorization that he or she makes under given conditions of illumination is very similar, if not identical, to the colour categorization (and experienced) by other individuals and hence that his or her experience has universal assent. In the sense that we use it here, “inherited”

implies a genetically inherited mechanism that engages the same neural connections and the same pattern of brain activity in all individuals, something which is applicable to the brain's ratio-taking system. The initial colour categorization, with the initial hue attached to it, (the two priors) are the consequence of that operation.

Our present paper is a step in a broader experimental enquiry, which aims to address the degree of variability produced in individuals when their experience is the product of interfacing the incoming visual signals with inherited and acquired brain concepts. The first study addressed the question of the biological basis of mathematical beauty (Zeki, Chén, & Romaya, 2018). Here we test the hypothesis that there will be high agreement among individuals of different ethnic and cultural backgrounds regarding the colour category into which they assign coloured patches when, regardless of their colour, the patches are made to reflect light of the same wavelength-energy composition. This may seem obvious but was important to demonstrate formally. To do so, we used the Land Mondrian experiment, where subjects view patches of different colour when each, in turn, is made to reflect the same triplet of energies, measured in terms of milliwatts per steradian per  $\text{m}^2$  ( $\text{mW.Sr}^{-1}.\text{m}^{-2}$ ). It is common knowledge that the colours of objects and surfaces do not change with fairly wide ranging changes in the wavelength-energy composition of the light in which they are viewed (Land & McCann, 1971; Land, 1986), a phenomenon generally referred to as colour constancy. We prefer to use the term “constant colour categorization” for two reasons; it avoids the use of the term “colour category” which in the past has been associated with use of language and because the hue (shade) of colour of a given surface or object does change with changes in the wavelength-energy composition of the light in which it is viewed, even if the colour categorization does not (Zeki *et al.*, 2017). We avoid the term colour category because, as used in past literature, it includes a lexical element; language is the element that we exclude in our studies, concentrating on non-linguistic matching instead. The distinction between *colour* and *colour categorization* is thus important for the experiments described here, in which we set out to learn the extent of variability in the experience of colour categorization between individuals without the use of language.

No one has determined the precise concept, in neural terms, which the brain uses to generate colours. But the ratio-taking formulations produced by Edwin Land and his colleagues (Land & McCann, 1971; Land, 1983; Land, 1986) are perhaps the easiest to use, given their mathematical precision and the predictability of the results produced through them. We acknowledge that many different ways of implementing this have been proposed (Foster, 2011) but they all share a common feature, namely a comparison of the wavelength composition of light reflected from different surfaces. This is what we, too, emphasize here although we rely more on the classical approach of Land and his colleagues, without implying that it is the final word on the implementation. The exact concept is in any case not critical for the work reported here but the experiments we have used and the avoidance of language are, and these are based on Land's.

In summary, we used the Land Mondrian experiments to investigate how subjects of different ethnic and cultural backgrounds categorize patches of different colour without the use of language, by matching the colours they experience when viewing different patches with that of a standard set of Munsell chips, since the use of linguistic criteria in studies of colour categorization (Berlin & Kay 1969/1991) has been criticized for not being equally applicable across languages (Biggam, 2012). We return to this in the Discussion.

## Material & Methods

*Subjects:* Twenty subjects, of whom 10 were females, took part in the experiment; their mean age was 24.1 years, with a standard deviation (s.d.) of 6.7. They were recruited through advertisements at University College London (UCL), were over 18 years of age and had normal or corrected to normal vision. They were all tested with Ishihara plates (Ishihara, 1988) for colour vision abnormalities and none was found to be deficient. No subject reported any neurological or psychological disorder, all gave informed consent and the experiment was approved by the UCL Ethics Committee (12327/001). Subjects came from the following countries: Cyprus, Thailand, Turkey, Lebanon, France, Ghana, China, Brazil, India, Greece, Egypt, Japan, Venezuela, Netherlands, United Kingdom and Spain; for all but one English was a second language.

*Task:* Subjects had to match the colour of patches in a Land Colour Mondrian display under specific conditions of illumination (see below) with coloured Munsell chips which were also viewed under specific illumination conditions. The patch to be matched was indicated to the subjects by a laser beam. In making the matches, subjects did not use words or language as their response. *The Mondrian displays:* As in Land's original experiments, we used Land Colour Mondrian displays, which were placed at a distance of 2 metres from the observers; they consisted of an assembly of squares and rectangles so arranged as to form an abstract scene with no recognizable shapes or objects, besides rectangles and squares. This controls for any effects due to memory and learning of what colours objects should have. To avoid specular reflectance, we used matt Color-Aid papers which reflect a constant amount of light in all directions. No patch was surrounded by another patch or patches of a single colour, thus avoiding induction effects.

We constructed and used two Mondrian displays, and subjects had to match the colours of each to Munsell chips (see Figure 1). Eight test patches were selected in each display, seven examples representing each of the basic colour categories Blue, Brown, Green, Yellow, Orange, Red, the "extra-spectral" Purple and one boundary colour – Turquoise - lying between Blue and Green. In the following description, each of the test patches will be referred to by the colour names given above although in the experiments subjects did not use language but merely matched the patch to the Munsell chips. Both Mondrian displays included the same eight test colour patches but in different configurations; in both, each patch subtended  $8.25^\circ \times 6^\circ$  and the surrounding patches extended more than  $10^\circ$  in all directions.

The Mondrian displays were illuminated by three carousel projectors (Kodak Ektagraphic B-2AR), equipped with ELH 120V 300W bulbs, rheostats and three gelatine filters passing long-, middle-, and short-wave light, respectively; the filters had been specially manufactured for Zeki's experiments by Edwin Land (Zeki, 1980). The long-wave filter transmitted light in the range of 592nm to the long end of the visible spectrum with a peak transmittance greater than 660nm. The transmittance of the middle-wave filter was in the range 492-580nm (peak 528nm) while the short-wave filter transmitted light in the range 386-493nm (peak 432 nm) with a secondary

peak at 700nm. Each projector was equipped with a separate rheostat and shutter, thus enabling the intensity of light coming from each to be adjusted separately.



*Figure 1. Appearance of Mondrian displays 1 (left) and 2 (right) under daylight viewing conditions. Color-Aid notations of test patches and corresponding names: 1. BW T2=Blue, 2. O S3=Brown, 3. YG T3=Green, 4. YO Hue=Orange, 5. V T1=Purple, 6. R Hue=Red, 7. GBG T4=Turquoise, 8. Primrose Yellow=Yellow.*

For each test stimulus, we adjusted the amount of long, middle and short-wave light of the three carousel projectors so that each patch, when judged for its colour, reflected a nearly constant ratio of 60% long-, 20% middle- and 20% short- wave light. Apart from the slightly different ratios used compared to the ones used in Land's experiments (Land, 1974; Land & McCann, 1971; Land, 1986), this constitutes a replication of Land's classical experiment in colour vision, the only difference here being that we were interested in the colour category, rather than the hue, to which matches were made. We note that our subjects never experienced the Color Aid patches in daylight conditions; hence their only acquaintance with our test stimuli was under the conditions described above. **The reason for selecting Color Aid matte patches for the construction of the Mondrian displays was because our three projectors – equipped with a long, medium and short wave filters – were illuminating the Mondrian displays from slightly different angles and we wanted each test surface to reflect, as much as possible, constant amount of light in all directions from all three projectors.** The energies reflected from each patch were

measured in milliwatts per steradian per square meter ( $\text{mW.Sr}^{-1}.\text{m}^{-2}$ ) separately for each projector using a PR-670 tele-spectroradiometer (Table 1). We also report the stimulus specifications for each test patch in  $10^\circ$  relative cone excitation units (Stockman & Sharpe, 2000) in Supplementary Table S1. The consistency of the ratios was checked before each experimental session.

*Table 1. Ratio of radiances ( $\text{mW/Sr}^{-1}/\text{m}^{-2}$ ) for long-, middle- and short wave light reflected from each test patch in the Mondrian displays.*

	Long wave light	Middle wave light	Short wave light
<b>Blue</b>	58.88	20.22	20.89
<b>Brown</b>	59.50	20.32	20.18
<b>Green</b>	61.05	18.81	20.14
<b>Orange</b>	59.85	20.12	20.03
<b>Purple</b>	60.41	19.43	20.16
<b>Red</b>	60.74	19.85	19.41
<b>Turquoise</b>	60.11	19.83	20.06
<b>Yellow</b>	60.11	20.06	19.83

### *The Munsell Chips*

The Munsell colour system was designed with the objective of representing perceptually uniform visual spacing of hue, Chroma (saturation) and Value (lightness) dimensions. In the viewing conditions specified above, subjects were asked to match the colour of the eight nominated patches with one of the 44 colour chips from the Munsell Book of Color (Glossy Collection, M40115). 40 hues of the Munsell set were selected to have the maximum available Chroma and variable Value levels to which we also added 4 darker stimuli for the yellow to red hues (Munsell 2.5YR to 10YR) because there were no brown or yellow chips at the same Value level (see Figure 2 and Supplementary Table S2 for the full specifications of the chips). The order of the chips was randomised and displayed on an annulus at a constant eccentricity of  $10^\circ$  from the point of fixation. They were presented against a mid-neutral grey surround inside a viewing booth illuminated by two GrafiLite daylight simulators (CIE 1931  $x = 0.327$ ,  $y = 0.339$ ). **In this study we were more interested in the constancy of colour categorization and we have chosen the glossy version of the Munsell system because it is the one from which chips used in most colour categorization studies (Berlin & Kay, 1969/1991; Sturges & Whitfield, 1995;**



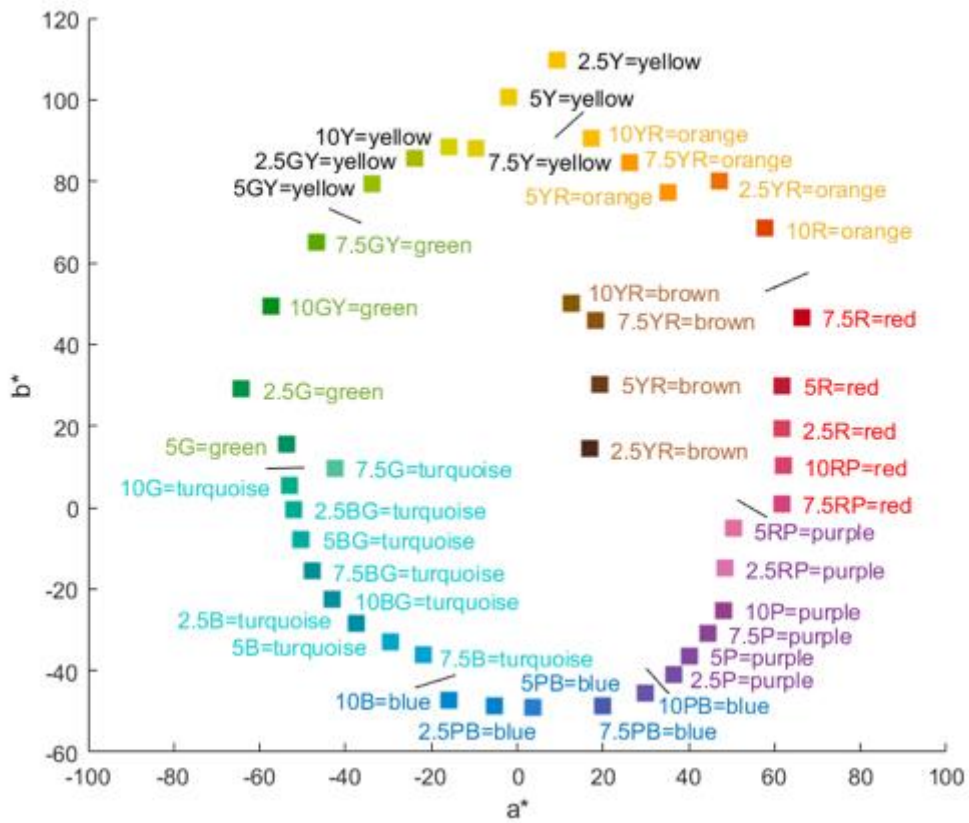
**Roberson *et al.*, 2005; Kay *et al.*, 2010) but in fact the use of the two would make little difference (Mylonas & MacDonald, 2010; Olkkonen *et al.*, 2010).**

### ***Colour Matching Procedure***

Participants had to match the colour of the eight test patches in each Mondrian display with the Munsell chip that, to them, was closest in colour to the patch under examination. The viewing booth with the Munsell chips was placed on a desk at a distance of 60cm from the observers. After adjusting the rheostats of the projectors to make each patch reflect the (same) amounts of long-, middle-, and short-wave light (given above), all three projectors were switched on to illuminate the entire Mondrian display, while the two daylight sources were switched on to illuminate the 44 Munsell chips. Any remaining light sources in the experimental room were eliminated. Participants performed the successive colour matching tasks without time limit but, in practice, each trial took less than 1 minute. The procedure was repeated twice for each test patch using the two different Mondrian displays of Figure 1 to measure the reliability of the responses, thus giving a total of 16 trials per subject.

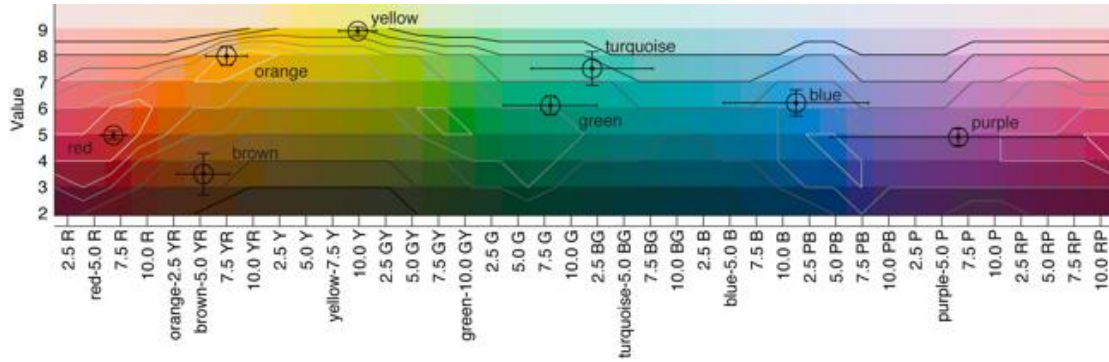
### ***Classification of hues into colour categories***

The exact hue or shade of a coloured surface varies under different conditions of illumination while its colour category remains constant (Zeki *et al.*, 2017). To classify the hues of the Munsell chips into categories without the use of language, based on the distance between them in colour space, we employed a k-means algorithm (which is reproducible using Matlab's k-means routine with default random settings) with Euclidean distances in CIELAB (D65), where the number of clusters was set equal to the number of the test samples ( $k=8$ ). In Figure 2, we show the Munsell chips of our comparison stimuli clustered into eight categories. This perceptual-based clustering corresponds well to the observed distribution of the colour terms on the surface of the Munsell system with the exception of pink (see supplementary Figure S1). For the sake of convenience during the analysis only, we supplemented this by assigning a colour name to the numerical outputs of the k-means algorithm (1-8), based on a colour naming estimator trained on the responses obtained from hundreds of English speaking participants in an online experiment (Mylonas & MacDonald, 2016) (Figure S1). We note that our subjects were naïve to this classification scheme.



**Figure 2.** Munsell chips shown as though viewed from above ( $a^*$   $b^*$  plane of CIELAB space) segregated into colour categories based on their in-between Euclidean distances.

The means and standard deviations of hue and Value of the matches (between the Munsell chips and each of the eight test patches) for all subjects and for both Mondrians are shown in Figure 3.



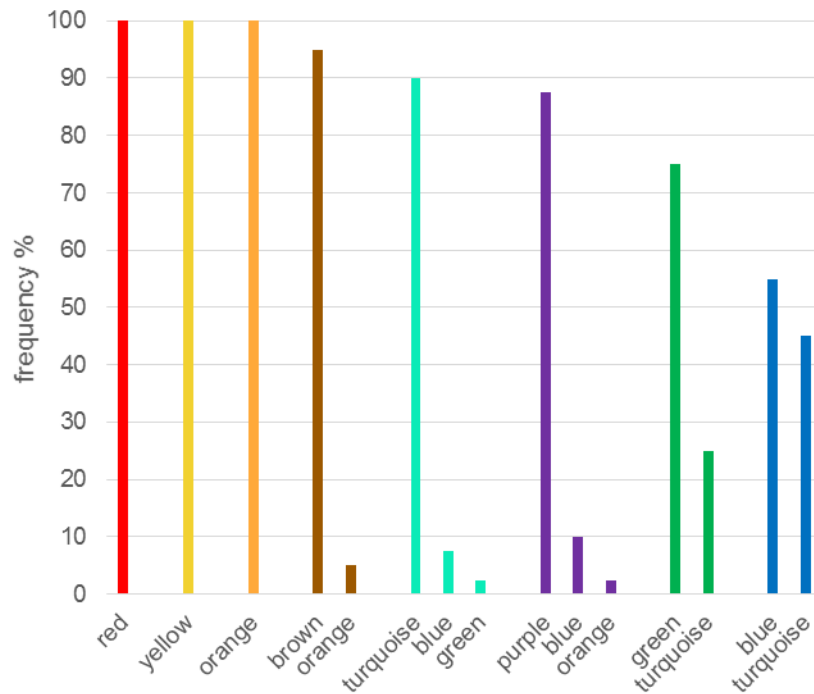
**Figure 3.** Variability of colour matching responses for each of the eight test patches for both Mondrian displays. Rings denote the mean and the error bars the standard deviation. The horizontal lines show the variability in the hue dimension and the vertical lines the variance in the Value dimension. Achromatic contours correspond to high levels (white), medium levels (grey) and low levels (black) of saturation (Chroma). The colour terms under each Munsell hue indicate the nearest hue of each test patch in terms of  $\Delta E_{00}$  in CIELAB.

The patches with the lowest hue variance in their matched Munsell chips were red, yellow, orange and brown, followed by green. The highest variability was observed for the ‘extra-spectral’ purple patch followed by blue and the boundary colour turquoise. Thus, the variability in colour matching responses is lower for reddish than bluish colours ( $r = 0.94$ ,  $n = 8$ ,  $p < 0.0005$ ); this reflects the smaller perceptual extents of categories in the warm region (in terms of steps leading to a change in hue), than in the cool region of colour space (Berlin & Kay, 1969/1991; Mylonas & MacDonald, 2016). A comparison of the means obtained for the different patches in the two Mondrian displays produced a good agreement (mean CIE  $\Delta E_{2000} = 2.00$ ), with the largest differences being observed for the brown and purple patches ( $\Delta E_{00} = 4.58$  and  $\Delta E_{00} = 2.23$ , respectively) and the smallest for the yellow patch ( $\Delta E_{00} = 1.08$ ). The uneven saturation levels of the Munsell surface (see achromatic contours) can only account for part of our results, namely for two (red and orange) out of eight test colours (Witzel & Franklin, 2014; Witzel, Cinotti & O’Regan, 2015).

The matching of the test patches showed that the categorization was constant, in other words that patches were neither identified as a different category nor was there a systematic shift towards the dominant long wavelength light.

The above description applies to hues of the Munsell chips; we were in fact more interested in the variability of colour categories, because it is the colour category rather than the hue that remains constant (Zeki *et al.*, 2017). In Figure 4, we convert the matches given in terms of the Munsell chips into categories following the k-means

classification scheme given above, i.e. without the use of language (Figure 2). There is no variability for matching the red, yellow and orange patches to their corresponding Munsell categories and high consistency for brown ( $\chi^2 = 61.2, p > 0.0001$ ), turquoise ( $\chi^2 = 48.05, p > 0.0001$ ), purple ( $\chi^2 = 42.05, p > 0.0001$ ), and green ( $\chi^2 = 18.05, p > 0.0001$ ) using Yate's correction for all chi-square tests. For the blue test patch, although the allocation was placed more frequently in the correct (blue) category, this allocation was not significant ( $\chi^2 = 0.45, p = 0.5$ ) as it was often placed within the turquoise category. It should be noted however that they were placed in the blue or turquoise (neighbouring) categories, in spite of the fact that they were reflecting twice the amount of long-wave compared to middle- or short-wave light.



**Figure 4.** Frequency of corresponding colour categories of Munsell chips selected as the best match for each test patch in both Mondrians. Colours denote the colour of the test patch and the labels show the corresponding clusters for the Munsell chips.

In summary, our results shows that, of a total of 320 responses, 88% allocated the test patches of the Mondrian displays to the Munsell chips belonging to the same category. The 12% were matched to chips belonging to other, but closely neighbouring, colour categories ( $\chi^2 = 250, p < 0.0001$ ). Of these, there was a high consistency of 90% in allocation of the boundary colour (turquoise) while the consistency was over 87% for 'extra-spectral' purple.

## Discussion

The experiments reported here constitute part of a series in which we explore, within the context of the brain as a knowledge-acquiring system, judgments that can reasonably be accounted for by supposing that they are based on biologically inherited concepts or mechanisms and are thus distinct from post-natally acquired ones (Zeki, 2009). Certain characteristics facilitate the categorization of experiences or judgments as being based predominantly or even exclusively on biologically inherited concepts. Prominent among these are: (a) a lesser variability between subjects, even those belonging to different races and cultures, when making judgments based on inherited concepts and (b) a resistance of such concepts to change through experience and learning (Zeki & Chen, 2019). The consequence of this more restricted variability is that the individual making a judgment based on inherited concepts is more entitled to assume that his or her judgment has universal validity and assent. This has so far been found to be true for aesthetic judgments of portraits and landscapes (Vessel, Maurer, Denker & Starr, 2018) as well as mathematical formulae experienced as beautiful (Zeki *et al.*, 2018), all of which we consider to belong to the biological category. Aesthetic judgments based on such concepts are characterized by lesser variability in judgment ratings compared to aesthetic judgments of man-made artefacts (which are more likely to be interfaced through synthetic concepts). In the work reported here, we extend this approach to colour vision and do so in a Bayesian context.

Berlin and Kay (1969/1991) proposed a total universal inventory of eleven basic colour categories (corresponding to the English terms: white, black, red, yellow, green and blue, grey, purple, brown, orange and pink) defined by a combination of multiple linguistic and psychological criteria. Such criteria have been strongly criticized as being not equally applicable across languages (Biggam, 2012). Others have identified basic colour categories on more objective behavioural criteria such as frequency, response time and consistency of colour naming responses (Boynton & Olson, 1987; Lindsey & Brown, 2014; Mylonas & MacDonald 2016). Recently, colour categories have been measured using simulated Munsell chips under varying lighting conditions; the strong correlation between naming consistency across illuminants and across

observers suggests a close link between categorical colour constancy and consistent colour communication (Olkkonen, Hansen & Gegenfurtner, 2009).

Because these approaches have been based on the use of language and because variations in categorizations are traceable to language, we opted for a different approach, one that is widely used to access colour constancy and offers comparable performance (Troost & De Weert, 1991; Speigle & Brainard, 1996); we investigated whether subjects of different linguistic and ethnic backgrounds categorize different colours in a similar way when light reflected from patches of different colour have the same wavelength-energy composition. We did so by asking subjects to match the experienced colours of the viewed patches with that of a standard set of Munsell chips, without the use of language.

### ***Colour categorization is dictated by inherited programs or concepts***

Colour is perhaps the most extreme example of an experience that is dictated by an inherited brain concept. We refer to this concept, based largely on the work of Edwin Land and his colleagues, as that of ratio-taking although one could equally refer to it as a brain algorithm or program (see Introduction). Specifically, the concept here is one in which light of any waveband reflected from a surface is compared with light of the same waveband reflected from surrounding surfaces, and a ratio between the two taken to categorize the colour ( *$\beta$  prior*) of the viewed patch; this constitutes the initial prior to which an initial hue is attached. Therefore the prior consists of two components – the categorization of the colour and the hue attached to that categorization. Although Land supposed that the ratio taking is applied three times, for long-, middle- and short-wave light, it is equally possible that it is done many times for lights of many different wavebands, **given the wide distribution in peak wavelength selectivity of cells in the cerebral cortex and particularly in area V4 (Zeki, 1980)**. The net result of these operations is that colour perception becomes largely independent of the continuous fluctuations in the wavelength-energy composition of the light reflected from a surface, thus leading to a perceptual stabilization of colours. It is common to suppose and write of the result of such a stabilization as colour constancy, by which is meant that the colours generated are constant and largely independent of the precise wavelength-energy composition of the light reflected off them. We believe, however, that describing the end result as a constant colour category is preferable because what does not change as a result of this

ratio-taking operation is in fact the colour category, not the hue (or shade of colour); the latter changes when surfaces are viewed in different illuminants (Zeki *et al.*, 2017).

Within a Bayesian context, of the two linked initial  $\beta$  priors in colour vision - the constant colour category and the hue attached to it - the latter (hue or shade of colour) can change with changes in wavelength composition of the light reflected from a patch or with changes in the arrangements of the surrounds, while the former (the colour category) either does not or changes much less. In this work, we showed that even when a patch (say the green or blue one) is reflecting more long-wave light, its colour category does not change to red or reddish colours but remains within the green category for all subjects. Our experiment is a first step in showing that there is a high consistency across individuals in matching, without the use of language, patches that differ in colour but reflect light of the same wavelength-energy composition, to definitive colour categories in the Munsell system. The implication of this is that the colour category  $\beta$  prior must be resistant to change or modification through experience and learning, in other words that there will be little variation in the posterior generated from the colour category  $\beta$  prior with experience and knowledge for individuals differing in their ethnic and cultural backgrounds. One would expect that to be so because the primordial function of the brain's colour system is to acquire knowledge about the world in terms of colour; it has to undertake this task when the signals and information reaching it from the outside world is never the same from moment to moment. Hence the brain must stabilize the world of colour as best it can, in order to be able to acquire knowledge about the constant and invariant properties of objects and surfaces in terms of colour. Hence, if a constant colour category biological prior were to be hostage to change through experience, then that would diminish considerably the ability of the brain's colour system to stabilize the world of colour and thus allow the organism to use colour as biological signalling mechanism and to communicate through it.

Both Helmholtz and Hering tried to account for colour constancy by invoking, in the case of Helmholtz higher cognitive factors such as learning and judgment (Helmholtz, 1911) and, in that of (Hering, 1877), "memory colours". Such factors may well play a role to modify the experience of colour for objects of which one has knowledge

through experience. But this, significantly, is not true for colour that is detached from definitive objects (Vandenbroucke, Fahrenfort, Meuwese, Scholte & Lamme, 2016) or colour attached to “nonsense” objects, of which Land’s experiments constitute a classic example and which is why we have used these experiments here. But even such knowledge about the colour of known objects may be over-ridden by the brain’s computational process to generate colours that make little sense in terms of the known colours of objects, as Land’s two colour projection experiments show (Land, 1959). In any case, the role of memory colours is controversial (Connolly, 2019); we agree with Connolly that, where it exists, it may enhance perceptual learning by making it easier to separate objects from their backgrounds and that it cannot be a mechanism for generating colour constancy as sometimes supposed (e.g. Olkkonen *et al.*, 2009) for the simple reason that the universality with which constant colour categorizations are generated in all humans makes it inconsistent with the supposition that memory colours enable colour constant categorization.

Consistent with the belief outlined above, what we have shown here, in summary, is that colour categories remain stable (Figure 4) and that such variation as there is, is rather in the hues within these constant colour categories (Figure 3). Thus of the two inter-connected  $\beta$  priors in colour vision, colour category and hue, the former remains constant, does not vary in terms of the posteriors produced from it when the wavelength-energy composition of the light reflected from it is varied, or when patches belonging to different colour categories are made to reflect the same wavelength-energy composition. The initial colour category  $\beta$  prior is thus highly resistant to modification through learning and experience. This represents a departure from the general Bayesian principle that priors lead to different posteriors and different beliefs attached to them through experience (Zeki & Chen, 2019). Specifically, a colour category can never become a posterior; it is always a prior. This is because, no matter what the wavelength-energy composition of the light reflected from, say, a green patch, it will always belong to the green category. Only the hues within that patch can become posteriors which can then act as priors for the generation of other (posterior) hues but ones which belong to the same colour category.



### ***Differences between ‘warm’ and ‘cool’ colours***

We undertook this study in the belief that the result of ratio-taking operations are similar in all humans. The consequence is that the results will also be similar in all humans, with little variability in the ascription of colours to given categories. The variability in matching the colour of the patches with chips belonging to different Munsell categories was indeed very limited, especially for red, yellow and orange (warm colours) while it was broadest for purple and blue (cool colours). The small variance in the red and yellow matches reflects the fact that, in terms of their extent measured as steps leading to a change in hue, they are indeed smaller categories than blue or purple, which have a larger number of hue steps and hence a higher variability (Berlin & Kay, 1969/1991; Mylonas & MacDonald, 2016).

This difference in the width between warm and cool colours, in terms of hue steps needed to change to another colour category, has been observed before. The work of (Gibson *et al.*, 2017) showed that warm colour categories are more salient than cool ones and (Danilova & Mollon, 2010), in their colour discrimination studies, showed a border between warm and cool categories, which we interpret as signifying that warm colours are narrower in terms of hue steps than cool ones. Although our results correspond to theirs in drawing a distinction between warm and cool colours, we do not interpret our results, as they do, in terms either of post-receptoral channels (Danilova & Mollon, 2010) or of communication needs (Gibson *et al.*, 2017). For us, the task was strictly a computational perceptual task and reflects the constancy of colour categories, reached by a computational brain process that is independent of learning, memory, environmental and social factors. And, while acknowledging that linguistic criteria may be of importance in classifying colours in terms of language, we believe that the classification according to colour categories is not dependent upon language and experience. Supporting evidence for this comes from the ability of children and monkeys to categorize colours much like adult humans (Sandell *et al.*, 1979; Skelton *et al.*, 2017).

It is worth adding that although we used “purple” as an example of an extra- spectral colour, one that is not considered to be a primary colour according to Hering (1877), the constancy of its categorization by different subjects was actually higher than that of blue and green, which are considered to be primary colours (see Figure 4).

Moreover, for turquoise which we used as an example of a boundary colour, and which is not considered to be either a primary colour or a basic colour term (see Berlin & Kay 1969/1991), was categorized with consistency, supporting earlier findings for its position as a basic colour (Mylonas & MacDonald, 2016). These findings suggest that colour categorizations involve higher cortical processes and cannot be considered only in terms of early perceptual mechanisms.

### *Limitations and future directions*

**It is worth emphasizing some limitations of the present study. Although we believe that our results are indicative of a fundamental uniformity, across races and cultures, in brain mechanisms that are engaged in categorizing colours, our sample nevertheless falls far short in that it was limited to 20 subjects. One future direction is to extend the sample significantly, not only by incorporating more subjects but by ensuring that there is an adequate representation of subjects from vastly under-represented populations. An interesting example is that of the Amazon Indian tribe, the Nambikwara, whose language does not distinguish between different shades of blue which is consequently often categorized with green or even black and for whom red and yellow “often” fall into the same category. Such a categorization that has been traced to the use members of the tribe make of seeds to prepare different dyes (Lévi-Strauss, 1955/2008). In addition, there might be subjects in industrialised societies, such as Russian speakers, who divide the unitary English blue category into light and dark segments (Paramei, Griber & Mylonas, 2018). This points to the importance, in future studies, of testing the constant categorization abilities of subjects from more diverse cultural, ethnic and environmental backgrounds.**

**Another limitation relates to the fact that we used mostly typical examples of colour categories and did not investigate the shape of the categories by including in our studies a large number of boundary colours. Our test boundary colour (turquoise) between blue and green revealed an underlying constant category important to the subjects which is not named using a traditional universal basic colour term (Berlin & Kay, 1969/1991) and thus might be independent of language. Our study thus says little about possible variabilities in the categorization of the borders of colour categories, even though it shows that**

**there is little variability when the matching is limited to more typical colours, which we believe to be an important step in demonstrating the restricted variability in categorizing colours. Future studies may well show that there is more variability in the shape of colour categories especially in the blue-green regions between subjects belonging to different races and different cultural and environmental backgrounds.**

**This ushers in another future direction that is important in pursuing studies in colour categorization. Although both monkeys and human infants appear to be able to categorize colours (see above), it is possible that subtler colour categories are acquired post-natally with exposure to different environments and that this fine-tuning may be dependent upon culture and environment. It would be fascinating to undertake such a study in the future, especially with humans who develop in environmentally conspicuously distinct surroundings.**

### *Conclusion*

In summary, there was a trivial variability in assigning colours that reflect a constant ratio of long-, middle- and short- wave light to different categories by subjects of different ethnic and cultural origins. This is a pointer to an important principle of the organization of the sensory brain, at least in terms of colour vision, namely that there is a very significant similarity in the inherited computational mechanisms for generating colour categories in all humans.

### **Competing interests**

We have no competing interests.

### **Authors' contributions**

SZ conceived and designed the study and authored the manuscript. AJ carried out the colorimetric measurements and collected the psychophysical data. DM coordinated the study, carried out the colour stimulus specifications and statistical analyses, and helped draft the manuscript. All authors gave approval for publication.

## Funding

DM was supported by the University College London (UCL) Computer Science-Engineering and Physical Sciences Research Council (EPSRC) Doctoral Training Grant: EP/M506448/1 - 1573073

## References

- Berlin, B and Kay, P. (1969/1991) *Basic Color Terms: Their Universality and Evolution*. University of California Press, Berkeley, CA.
- Biggam, C.P. (2012) *The Semantic of Colour: A Historical Approach*. Cambridge University Press, Cambridge, UK.
- Boynton, R.M and Olson, C.X. (1987) Locating basic colors in the OSA space. *Color Res. Appl.*, **12**, 94–105.
- Connolly, K. (2019) *Perceptual Learning*. Oxford University Press, Oxford and New York.
- Danilova, MV and Mollon, J. (2010) Parafoveal color discrimination: A chromaticity locus of enhanced discrimination. *J. Vis.*, **10**, 1–9.
- Dayan, P., Hinton, G.E.E., Neal, R.M.M., & Zemel, R.S.S. (1995) The Helmholtz machine. *Neural Comput.*, **7**, 889–904.
- Foster, D.H. (2011) Color constancy. *Vision Res.*, **51**, 674–700.
- Friston, K., Mattout, J., & Kilner, J. (2011) Action understanding and active inference. *Biol. Cybern.*, **104**, 137–160.
- Gibson, E., Futrell, R., Jara-Ettinger, J., Mahowald, K., Bergen, L., Ratnasingam, S., Gibson, M., Piantadosi, T. S., Conway, B.R. (2017) Color naming across languages reflects color use. *Proc. Natl. Acad. Sci. USA*, **114**, 10785–10790.
- Helmholtz, H. von (1911) *Handbuch Der Physiologischen Optik*. Leopold Voss, Hamburg.

- Hering, E. (1877) *Outlines of a Theory of the Light Sense*. Harvard University Press, Cambridge, MA.
- Kant, I. (1781) *Kritik Der Reinen Vernunft, 1st Edition, Translated by WS Pluhar (1996) as Critique of Pure Reason*, 1st edn. Hackett, Indianapolis.
- Kay, P., Berlin, B., Maffi, L., Merrifield, W., & Cook, R. (2010). *The World Color Survey*. Berkley, CA: CSLI Publications.**
- Land, E. (1974) The retinex theory of colour vision. *Proc. R. Inst. G. B.*, **47**, 23–58.
- Land, E.H. (1959) Color vision and the natural image. Part 1. *Proc. Natl. Acad. Sci. U. S. A.*, **45**, 115–129.
- Land, E.H. (1983) Recent advances in retinex theory and some implications for cortical computations: color vision and the natural image. *Proc. Natl. Acad. Sci. U. S. A.*, **80**, 5163–5169.
- Land, E.H. (1986) An alternative technique for the computation of the designator in the retinex theory of color vision. *Proc. Natl. Acad. Sci. U. S. A.*, **83**, 3078–3080.
- Land, E.H. & McCann, J.J. (1971) Lightness and Retinex Theory. *J. Opt. Soc. Am.*, **61**, 1–11.
- Lévi-Strauss, C. (1955/2008). *Tristes tropiques*, pp 3-445 in in Œuvres, Bibliothèque de La Pléiade, Gallimard, Paris.**
- Lindsey, D. T., & Brown, A.M. (2014) The color lexicon of American English. *J. Vis.*, **14**, 1–25.
- López, J.D., Litvak, V., Espinosa, J.J., Friston, K., & Barnes, G.R. (2014) Algorithmic procedures for Bayesian MEG/EEG source reconstruction in SPM. *Neuroimage*, **84**, 476–487.
- 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 and MacDonald, L. (2016) Augmenting basic colour terms in English. *Col. Res. Appl.*, **41**, 32–42.
- Olkkonen, M., Hansen, T., & Gegenfurtner, K.R. (2009) Categorical color constancy for simulated surfaces. *J. Vis.*, **9**, 6.
- Olkkonen, M., Witzel, C., Hansen, T., & Gegenfurtner, K. R. (2010). Categorical color constancy for real surfaces. *Journal of Vision*, 10(9).  
<https://doi.org/10.1167/10.9.16>**
- 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. <https://doi.org/10.1002/col.22190>**
- Pouget, A., Beck, J.M., Ma, W.J., & Latham, P.E. (2013) Probabilistic brains: knowns and unknowns. *Nat. Neurosci.*, **16**, 1170–1178.
- Rao, R.P. & Ballard, D.H. (1999a) Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci.*, **2**, 79–87.
- Rao, R.P.N. & Ballard, D.H. (1999b) Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci.*, **2**, 79–87.
- Roberson, D., Davidoff, J., Davies, I. R. L., & Shapiro, L. R. (2005). Color categories: Evidence for the cultural relativity hypothesis. *Cognitive Psychology*, 50(4), 378–411. <https://doi.org/10.1016/j.cogpsych.2004.10.001>**
- Sandell, JH, Gross, CG, and Bornstein, M. (1979) Color categories in macaques. *J. Comp. Physiol. Psychol.*, **93**, 626–635.

- Skelton, A. E., Catchpole, G., Abbott, J. T., Bosten, J. M., & Franklin, A. (2017) Biological origins of color categorization. *Proc. Natl. Acad. Sci.*, **114**, 5545–5550.
- Stockman, A., & Sharpe, L.T. (2000) The spectral sensitivities of the middle- and long-wavelength-sensitive cones derived from measurements in observers of known genotype. *Vis. Res*, **40**, 1711–1737.
- Sturges, J., & Whitfield, A. (1995). Locating basic colours in the Munsell Space. *Color Research and Application, Volume 20*(Number 6), 364–376.**
- Troost, J. M., & De Weert, C.M.M. (1991) Naming versus matching in color constancy. *Percept. Psychophys.*, **50**, 591–602.
- Vandenbroucke, A.R.E., Fahrenfort, J.J., Meuwese, J.D.I., Scholte, H.S. and Lamme, V.A.F. (2016) Prior knowledge about objects determines neural color representation in human visual cortex. *Cereb. Cortex*, **26**, 1401–1408.
- Vessel, EA, Maurer, N, Denker, AH, & Starr, G. (2018) Stronger shared taste for natural aesthetic domains than for artifacts of human culture. *Cognition*, **179**, 121–131.
- 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., Cinotti, F., & O’Regan, J. K. (2015). What determines the relationship between color naming, unique hues, and sensory singularities: Illuminations, surfaces, or photoreceptors? *Journal of Vision*, *15*(8), 19.  
<https://doi.org/10.1167/15.8.19>
- Zeki, S. and Chen, O. (2019) The Bayesian-Laplacian Brain. *Eur. J. Neurosci. in the press*.

Zeki, S. Chén, O.Y. and Romaya, J.P. (2018) The Biological Basis of Mathematical Beauty. *Front. Hum. Neurosci.*,

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

Zeki, S. (2009) *Splendors and Miseries of the Brain: Love, Creativity and the Quest for Human Happiness*. Wiley-Blackwell, Oxford.

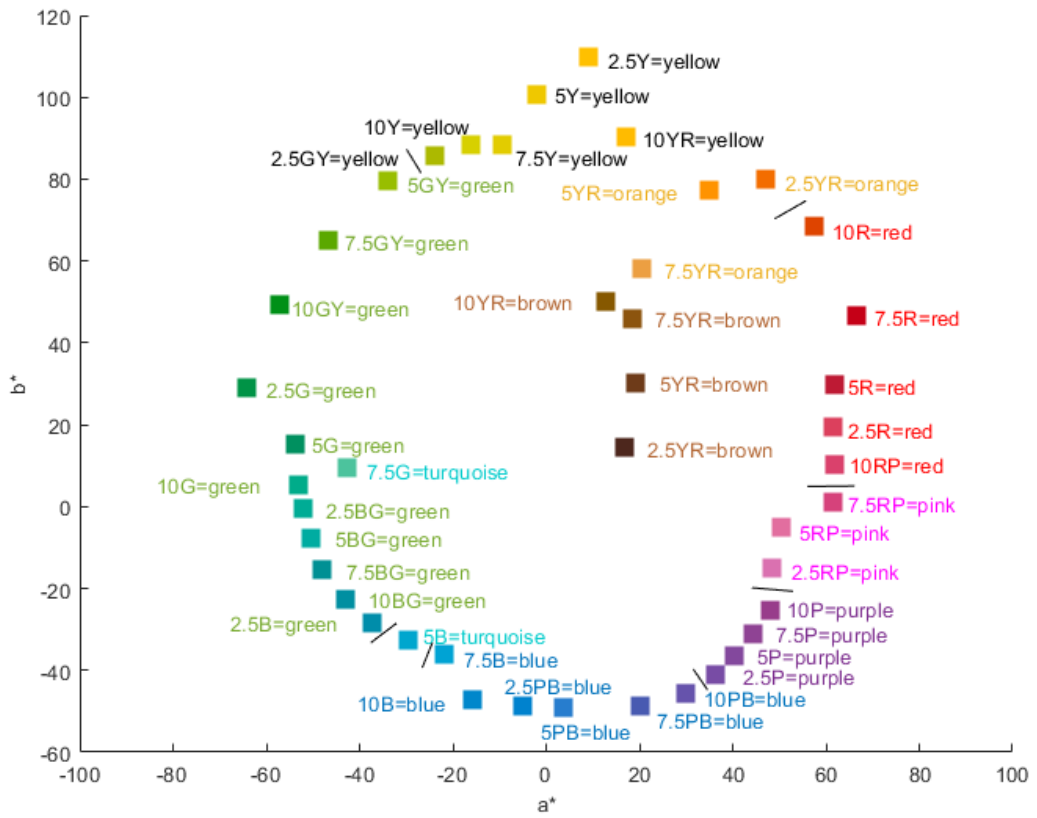
Zeki, S., Cheadle, S., Pepper, J., & Mylonas, D. (2017) The constancy of colored after-images. *Front. Hum. Neurosci.*, **11**, 10.338.

## Supplementary Material

*Table S1. Long (L), Medium (M) and Short (S) cone excitation ratios (Stockman & Sharpe, 2000) in 10-degrees for each of the eight test stimuli of the Mondrian displays.*

	<b>L</b>	<b>M</b>	<b>S</b>
<b>Blue</b>	0.45	0.31	0.23
<b>Brown</b>	0.47	0.31	0.22
<b>Green</b>	0.45	0.32	0.23
<b>Orange</b>	0.47	0.31	0.22
<b>Purple</b>	0.46	0.31	0.23
<b>Red</b>	0.47	0.32	0.21
<b>Turquoise</b>	0.45	0.32	0.23
<b>Yellow</b>	0.46	0.32	0.22





**Figure S1.** Munsell chips shown as though viewed from above ( $a^*$   $b^*$  plane of CIELAB space) segregated into lexical colour categories by a colour naming model trained by human observers (Mylonas & MacDonald, 2016).

**Table S2.** Munsell notation of comparison stimuli.

N°	Hue	Value	Chroma
1	2.5R	5	14
2	5R	4	14
3	7.5R	4	16
4	10R	5	16
5	2.5YR	6	16
6	2.5YR	2	4
7	5YR	7	14
8	5YR	3	6
9	7.5YR	7	14
10	7.5YR	4	8
11	10YR	8	14
12	10YR	4	8
13	2.5Y	8	16
14	5Y	8	14
15	7.5Y	8	12
16	10Y	8	12
17	2.5GY	7	12
18	5GY	7	12

19	7.5GY	6	12
20	10GY	5	12
21	2.5G	5	12
22	5G	5	10
23	7.5G	7	8
24	10G	6	10
25	2.5BG	6	10
26	5BG	6	10
27	7.5BG	5	10
28	10BG	5	10
29	2.5B	5	10
30	5B	6	10
31	7.5B	6	10
32	10B	5	12
33	2.5PB	5	12
34	5PB	5	12
35	7.5PB	4	12
36	10PB	4	12
37	2.5P	4	12
38	5P	4	12
39	7.5P	4	12
40	10P	4	12
41	2.5RP	6	12
42	5RP	6	12
43	7.5RP	5	14
44	10RP	5	14

---