This electronic thesis or dissertation has been downloaded from Explore Bristol Research, http://research-information.bristol.ac.uk

Author:
Muchhala, Mubaraka

Title:
Does the geometry of perceptual colour space reflect the colours in our environment?

General rights
Access to the thesis is subject to the Creative Commons Attribution - NonCommercial-No Derivatives 4.0 International Public License. A copy of this may be found at https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode This license sets out your rights and the restrictions that apply to your access to the thesis so it is important you read this before proceeding.

Take down policy
Some pages of this thesis may have been removed for copyright restrictions prior to having it been deposited in Explore Bristol Research. However, if you have discovered material within the thesis that you consider to be unlawful e.g. breaches of copyright (either yours or that of a third party) or any other law, including but not limited to those relating to patent, trademark, confidentiality, data protection, obscenity, defamation, libel, then please contact collections-metadata@bristol.ac.uk and include the following information in your message:

- Your contact details
- Bibliographic details for the item, including a URL
- An outline nature of the complaint

Your claim will be investigated and, where appropriate, the item in question will be removed from public view as soon as possible.

Does the geometry of perceptual colour space reflect the colours in our environment?

Mubaraka Muchhala

A dissertation submitted to the University of Bristol in accordance with the requirements for award of the degree of Doctor of Philosophy in the Faculty of Life Sciences

School of Psychological Science

April 2022

Word count: 42,032

Abstract

There are universal patterns in colour naming (Berlin & Kay, 1991; Cook et al., 2005) and the colour categories in a language influence colour perception during perceptual uncertainty (Bae et al., 2015). This suggests that that colour categorisation is not structured at random, but rather the structure is constrained to serve an adaptive purpose. Previous research has focused on identifying constraints on language and the human visual system to explain the origins of universal colour categories, but there was little evidence demonstrating ecological constraints on colour categories.

This thesis proposes that the distribution of colours across objects in the environment place an ecological constraint on colour perception and categorisation. If there are systematic patterns in colouration across behaviourally relevant objects, then colour categories may be formed to guide perceptual processes towards visual stimuli which are the most important. To explore this, two tasks were used: a colour estimation task to measure the geometry of perceptual colour space, and an object recognition task to measure the distribution of colours across objects.

Humans exhibited categorical biases in colour perception across hue and saturation towards universal colour categories, corresponding to red, blue, green, pink, orange and grey. These categorical biases were successfully estimated from the statistical regularities of colours across objects for two independent image datasets: colour perception was biased towards category foci which were most informative about objects, and away from category boundaries which were least informative about objects. These findings support a universal mechanism through which sensory systems adapt to environmental statistics, allowing for optimal representation of sensory features under uncertainty for both perception and language.

Acknowledgements

I'd like to express my deepest gratitude to Prof. Nick Scott-Samuel and Dr. Roland Baddeley for their unwavering guidance and support. Seven years ago, Nick inspired my passion for colour perception, and ever since he has encouraged and enabled me to keep exploring it. In addition, it is Roland's enthusiasm that has driven me forward whilst he also made sure I was stepping back. You both make a great team, and it was a privilege to receive your supervision.

I'd also like to thank Dr. John Fennel for the many times that he helped me with programming colour spaces. I am very grateful for the moral support of my PhD friends, Daria, Milton, Greig, Katie and Laura, who's daily company I greatly missed during the pandemic. And thank you so much to Anne who held me accountable, motivated me and added a bit of sunshine to my day during the write up.

A final thank you to my parents and brother who have always supported and believed in me. And Mayowa, Kim and Yasin for listening to me, encouraging me and providing welcome distractions. It's difficult to imagine completing this thesis without your support.

Declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED: .... DATE: 07/04/2022

Table of Contents

List of Figures

List of Tables

List of Equations

Abbreviations

LGN – Lateral geniculate nucleus

JND – just noticeable difference

BK – Berlin & Kay color naming data

BCT – basic colour term

WCS - World Color Survey

CCP - Categorical colour perception

KDE – kernel density estimation

1 Literature Review

1.1 Colour vision

Colour vision is the multi-stage process by which the visible portion of the electromagnetic spectrum is detected in the retina, and transformed into neural signals which allow humans to discriminate between colours in the environment. The stages of vision are hierarchical, with visual inputs being successively processed along the ventral stream to achieve visual recognition. Early stages are referred to as low-level vision, where visual input is processed to analyse local structure (Marr, 1982). Later stages are referred to as high-level vision, where visual input is matched to visual memory to facilitate image recognition and extract behaviourally relevant information (Marr, 1982).

When light enters the eye, cone photoreceptors in the retina are sensitive to different wavelengths of the light. There are three types of cones, which are sensitive to long, medium and short wavelengths (Masland, 2001; Stockman & Sharpe, 2000). Retinal ganglion cells in the optic nerve transmit signals from the cones to the lateral geniculate nucleus (LGN) in the thalamus, where the relative cone excitations are compared in three opponent pairs (De Valois & De Valois, 2001; Livingstone & Hubel, 1987). An achromatic dark-light channel adds long and medium wavelengths(L+M), a red-green channel contrasts long and medium wavelengths (L-M); and a blue-yellow channel contrasts short wavelengths with a combination of long and medium wavelengths (S-(L+M)). The comparison of photoreceptor signals is what enables humans to decode spectral information from the cones. Opponent signals from the LGN are transmitted into V1, V2 and V4 in the visual cortex (Zeki et al., 1991), where the three opponent channels are combined to analyse local image structure and produce a stable perception of colour.

Colour is perceptually organised in three dimensions: hue, saturation and brightness. Colour appearance reflects the opponent processes within colour vision, where a mixture of opponent colours cannot be perceived (e.g. reddish-green), but non-opponent colours can be perceived (e.g. reddish-yellow is orange, and greenish-blue is turquoise). Colour perception can be measured using just noticeable differences (JNDs) between colours. Participants are presented with two colours on two halves of a circle using a mixture of red, green and blue lights. One colour is adjusted using the intensity of each light, until there is a JND between the colours. The differences can be plotted to create a colour appearance model. These models are perceptually uniform, where the distance between colours in the model is reflective of the perceived difference between colours.

It was traditionally believed that only achromatic signals are used to process information in low-level vision, such as identifying edges to segment images into objects and scenes (Livingstone & Hubel, 1987). However, more recent evidence demonstrates that spatially selective neurons in V1 are also sensitive to colour (Johnson et al., 2001). In addition, trichromatic colour vision provides a behavioural advantage by improving the detection of ripe fruits among foliage (Regan et al., 2001). This suggests that colours (in particular hues along the red-green axis) are useful for segmentation for certain behaviourally relevant tasks. For over 700 images of natural scenes, Hansen and Gegenfurtner (2009) found that chromatic edge contrast and luminance edge contrast were independent but equally useful sources of information about edges, which could be combined to segment images. Hansen and Gegenfurtner (2017) also compared human labelled edges in scenes with luminance and chromatic edges in scenes, and found that colour information was a better predictor of human labelled edges than luminance edges, although both types of edge are useful. Altogether this evidence suggests that colours are used to analyse visual form and segment images into objects and scenes during low level visual processing, and therefore that colour is important for early visual processes.

Object recognition occurs within high-level vision, where visual input is matched to visual memory to extract meaning from the image. It was traditionally argued that objects are recognised using only shape information, where a structural description of the object is matched to structural representations of the object, in order to identify the object's name and meaning (Biederman, 1987). However, more recent evidence demonstrates that colour is also useful for high-level vision, such as recognition of objects, scenes or faces.

Many studies have observed that coloured objects are easier to recognise than black and white objects (Davidoff & Ostergaard, 1988; Laws & Hunter, 2006; Price & Humphreys, 1989; Tanaka & Presnell, 1999; Wurm et al., 1993). In a meta-analysis of 35 experiments comparing object recognition for regular versus greyscale images of objects, Bramão et al. (2011) found strong evidence that colour information improves object recognition. Additionally, the colour of pixels in fruits and vegetables fall within a very narrow distribution of hues in CIELAB space (Ennis et al., 2018), suggesting that hue can define a fruit or vegetable.

Furthermore, retrieving information about object colour activates the same areas of the brain that are involved in object recognition. Zeki and Marini (1998) displayed naturally and unnaturally coloured images of objects to participants and compared the brain activity between conditions using fMRI. They observed that viewing a naturally coloured image activated the same brain areas that are activated during tasks when participants retrieve object colour (fusiform gyrus, hippocampus and ventrolateral portion of the frontal cortex), and that these areas were not activated when viewing unnaturally coloured images. This suggests that colour is useful for the retrieval of information about objects, possibly because colour is an additional cue to aid retrieval of the object from memory.

Shape is obviously still very important for object recognition, as colour alone is not informative enough about object identity. However, colour information is very useful and should not be disregarded, particularly in instances where shape information is uncertain. For example when the object is occluded by another object, colour information is useful to reduce uncertainty (Tanaka et al., 2001). It is likely that colour and shape information interact in order for humans to recognise objects. Altogether, it is clear that colours are useful to help humans identify objects, therefore the colours which are most behaviourally relevant is likely to depend on the colours of objects in the environment.

1.2 Colour naming

Humans are able to perceive over two million distinguishable colours (Linhares et al., 2008; Pointer & Attridge, 1998) which vary on a continuum. All cultures must assign lexical labels to the continuous colour space to communicate about the most behaviourally relevant colours. Colour lexicons need to contain enough colour terms that information about colour can be communicated precisely, whilst not having too many terms which are difficult to remember and use regularly.

To measure how different languages solve this problem, Berlin and Kay (1991) collected data from 20 different languages using a colour naming task (this dataset will be referred to as BK). Participants were first shown an array of colours from the Munsell colour chart, and asked to verbally recall the colour in their language. Participants then identified the focal point and outer boundary for each colour term used. Berlin and Kay (1991) also looked at dictionary data from an additional 88 languages. A basic colour term (BCT) was defined as a colour which: 1) is monolexemic, such that its meaning cannot be deduced from the meaning of its parts, 2) has no overlap in meaning with another term, 3) can be used to describe a wide range of objects of the appropriate colour, and 4) is reliably used by most native speakers and across different occasions of use.

When comparing results across languages, Berlin and Kay (1991) observed that the focal points of colour categories formed universal clusters across the Munsell chart. They discovered 11 BCTs which are observed across languages universally: white, black, red, green, yellow, blue, brown, purple, pink, orange and grey. Languages varied in the number of BCTs used. As languages become more advanced, they develop new colour terms in a sequential order through seven stages (Figure 1.1). Thus, the BCTs in a language can be predicted by the number of colour terms in the language. If a language contains a certain colour term (e.g. blue) then it will also contain the colour terms from all previous stages (e.g. white, black, red, green and yellow). Given that languages share broadly similar intuitions about colour naming, Berlin and Kay (1991) advocated for a universal theory of colour categorisation.

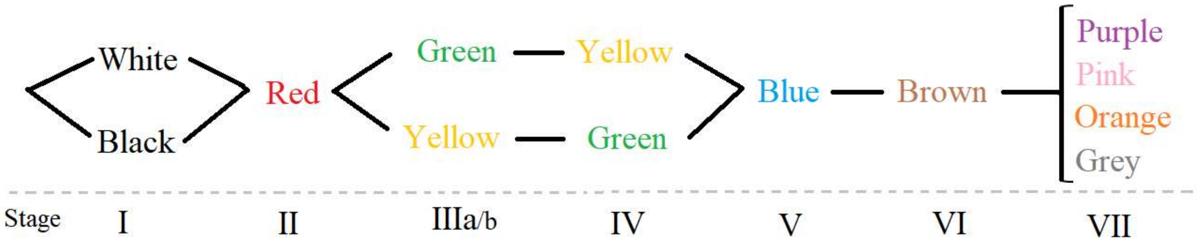


Figure 1.1 *Evolutionary sequence of BCTs. Each progressive stage represents the colour terms adopted by a language as the language becomes more advanced.*

These observations of universal patterns in colour naming were highly influential, however there were some issues with the experimental design. For some languages, there was only one participant per language. There is significant variability in colour naming among speakers of the same language (Kuehni, 2001; Lindsey & Brown, 2009), therefore one participant per language is not enough to estimate colour naming within that language. In addition, all the participants were English bilinguals living in America, therefore all of the languages studied were from industrialised societies. Berlin and Kay (1991) were criticised for taking a very Eurocentric viewpoint by endorsing evolutionary progress towards the English language and disregarding anomalous cases of colour term use in East Asian languages (such as Cantonese, Mandarin, Japanese and Vietnamese).

To evaluate the findings by Berlin and Kay (1991), Cook et al. (2005) conducted the World Color Survey (WCS), where they collected colour naming data from languages of 110 non-industrialised societies, with an average of 24 participants per language. Participants were successively shown 330 Munsell colour chips, and asked to name the colour in their native language. After naming all of the chips, the experimenter selected the BCTs in the language according to the BCT criteria, and participants identified the example colour chip for each BCT. The best examples for each BCT in non-industrialised societies clustered near the category foci for 6 English BCTs: white, black, red, green, yellow and blue (Kay & Regier, 2003; Lindsey & Brown, 2006, 2009; Regier et al., 2005). Therefore, similar patterns in colour naming were identified in industrialised and non-industrialised languages, suggesting that these patterns in categorisation are universal.

Kay and Regier (2003) investigated the colour naming data from the WCS and BK. For each language, they averaged colour naming data across speakers, and represented the colour terms as centroids in CIELAB space. Dispersion of colour terms was measured using Monte Carlo methods, where the distance between similar colour terms was summed across languages. The WCS naming data were significantly less dispersed than dispersion by chance, which suggests that colour terms are clustered close together across languages. In addition, the naming data from WCS were much more similar to the BK naming data than a randomly rotated dataset, which suggests that there is a high level of similarity in colour naming between industrialised and non-industrialised languages.

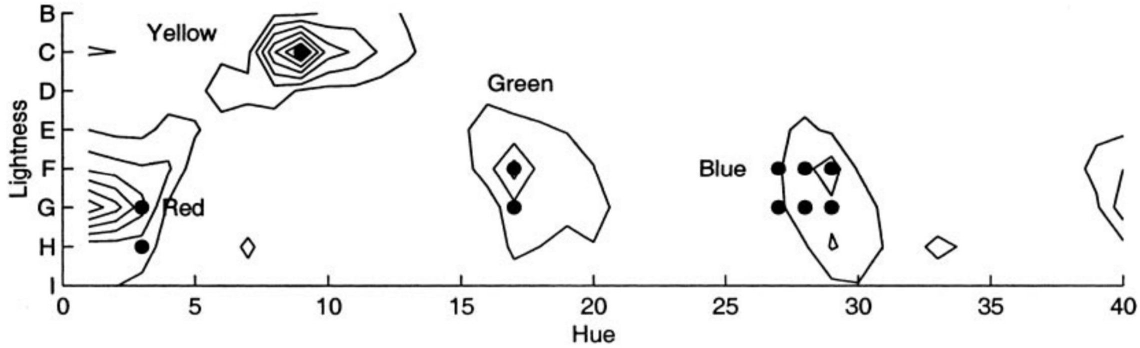


Figure 1.2 *Contour plot showing frequency with which colour chips were selected as category foci in the WCS for English language. Dots show best examples of English colour terms from BK (there are multiple best examples for each BCT). Reprinted from "Focal colors are universal after all." By T. Regier, P. Kay and R. S. Cook, 2005, Proceedings of the National Academy of Sciences, 102(23), p. 8386-8391. Copyright 2005 by The National Academy of Sciences of the USA.*

Regier et al. (2005) also looked at the focus data in the WCS (where participants identified the focal colour chip for each colour term). A contour plot showed the frequency with which colour chips were selected as category foci across all languages and speakers. The best examples of colour categories identified in the WCS cluster near the best examples of white, black, red, green, yellow and blue in BK (Figure 1.2). This suggests that there are universal focal colours which anchor colour lexicons universally. In addition, the WCS focus data were closer to the BK focus data than to the centre of the centroids from the WCS naming data. Therefore category foci aren't just a secondary feature that result from partitioning the colour space, but rather represent some universal anchoring points within the colour space which constrain colour naming.

In the above studies, colour naming data were averaged across speakers within a language. Lindsey and Brown (2006) took a different approach by applying k-means clustering on individual-level data across all languages (2367 informants). As k (the number of clusters) increased from 2-10, the colour categories revealed by clustering analysis progressed through stages similar to the Berlin and Kay evolutionary sequence. When k=2, the categories were split into warm (red, yellow, orange and pink) and cool (blue and green). As more clusters were added, the model produced fine grained colour categories reminiscent of English BCTs.

Lindsey and Brown (2009) proposed that the idiolects of WCS participants fall into 3-6 universal colour naming systems, which include a selection of the 11 BCTs. These naming systems occur universally with individual variation both within and across languages, such that multiple colour naming systems may be used in the same language. They therefore concluded that colour naming is not entirely determined by culture, as there is variation within cultures.

Altogether, the outcome of cross-cultural colour naming research lends strong support for universal patterns in colour naming, which are anchored by focal points in the colour space. Although there is variation in colour naming systems both within and across languages, there are constraints on the variation in both industrialised and non-industrialised cultures.

However, the WCS did receive some heavy criticisms, in particular regarding the stimuli used to test colour naming, and the criteria used to determine what constitutes a BCT

(Saunders & van Brakel, 1997). The stimuli used to test colour naming were 330 Munsell colour chips, which varied across hue and lightness, with maximum saturation. Maximum saturation varies strongly across the Munsell space, therefore may confound colour naming data.

Witzel et al. (2015) observed that focal colour choices correlate with saturation of Munsell chips, where local peaks in saturation were close to the foci for red, yellow, green and blue. This suggests that saturated colours are more likely to be selected as focal, and could explain clustering of focal colours across languages. Additionally, measuring subjective sensitivity to saturation using JNDs in CIELUV colour space, there is no evidence to suggest that subjective saturation is higher for focal colours in red, yellow, green or blue (Witzel et al., 2019; Witzel & Franklin, 2014). Therefore, the local peaks in saturation in Munsell space and their effect on colour naming in the WCS do not represent a universal property of colour perception but are rather a feature of the Munsell space. It is possible that participants in the WCS selected the peaks in saturation as focal points of chromatic colour categories as they were the most different from achromatic colours in Munsell space.

Jameson and D'Andrade (1997) proposed that if a colour space is irregularly shaped, then there is an optimal method of partitioning the space into categories, to maximise the informativeness of the colour lexicon. An optimal partitioning of an irregularly shaped space will maximise similarity of colours within categories and minimise similarity of colours across categories. To test whether universal patterns in colour naming reflect optimal partitioning of colour space, Regier et al. (2007) converted the colour chips used in the WCS into CIELAB colour space to measure the perceptual distances between colours. A mode map was created for each language, where each colour chip was labelled as its most commonly used colour term. When plotted in CIELAB space, the Munsell colour chips form a distorted sphere, with a pole in the centre of the sphere ranging from black to white, and chromatic chips arranged around the pole in a bumpy sphere (See Figure 2 in Regier et al., 2007).

Well-formedness was used to measure the optimality of a colour naming system, which is the sum of similarity between colours within a category, plus the sum of dissimilarity (1-similarity) between colour chips in different categories. The higher the well-formedness, the more optimal the partitioning of the space. Regier et al. (2007)

constructed simulations of the theoretically optimal partitioning of this space for colour lexicons with n=3-6 terms, by randomly assigning chips to n categories, and selecting the simulated lexicon with the highest well-formedness. The mode maps were similar to the theoretically optimal configuration in some languages, but many languages were also very different. This suggests participants in the WCS were not categorising Munsell colour chips based on irregularities in the perceptual colour space, and therefore distortions in Munsell chroma cannot alone explain universal patterns in colour naming.

The second major criticism of both BK and WCS are that the task instructions were complex and restrictive. First of all, the criteria for BCTs to be monolexemic and to not have any overlap in meaning with other terms seem clear in English, but may not be easily applied in non-English languages (Witzel, 2019). For example, only 6 BCTs where identified in Mandarin Chinese by Berlin and Kay (1991); however Gao and Sutrop (2014) identified 9 BCTs in Mandarin based on the cognitive salience of colour terms (i.e. how commonly the terms were used by speakers). Gao and Sutrop (2014) argued that there are BCTs in Mandarin which are integral to the language and should not be excluded, despite the terms containing multiple characters (therefore challenging the monolexemic rule).

In the guidelines for defining a BCT, secondary rules also discouraged colour terms which are recent foreign loan words or homonyms (where the colour term is also the name of an object which typically has that colour, such as gold). However, research in linguistic anthropology shows that colour terms commonly originate from object names or are borrowed from other languages (Biggam, 2012). Colour terms which were recently acquired in a language are easier to identify as foreign loan words or homonyms. For example, the indigenous Dani society was notably identified as the only example of a language with two colour terms (Berlin & Kay, 1991; Heider & Olivier, 1972). However, more recently Groh (2016) conducted an independent study of the Dani colour naming system, and identified 10 colour terms which were commonly used by Dani speakers and therefore argued to be important colour terms in the language, despite seven of the terms being object-related homonyms.

The uncertainty around what constitutes a BCT leaves much room for variance across languages. In the WCS, instructions to field workers outlined the definitions of BCTs as found in BK, with example colour terms given in English. Field workers asked participants to name the chips, and then determined a list of BCTs for that language based on the findings and instructions. This list of BCTs was then used to collect data on focal colours. Therefore the instructions could have resulted in the field workers implementing bias during the interviews, such as coaxing specific types of answers from participants based on their expectations from English language.

Even in American English, using less strict criteria for colour naming results in more than 11 high consensus colour terms, such as peach, teal, lavender and maroon (Lindsey & Brown, 2014). This suggests that the colour lexicon of American English is continuing to evolve, and therefore further supports the idea that there is no clear universal dichotomy between basic and non-BCTs. Future research may benefit from considering the basicness of colour terms on a gradual spectrum based on variables such as consensus or frequency of use. Despite issues with the experimental design in cross-cultural approaches, it is clear that colour categorisation does not randomly vary across languages: universal patterns exist.

1.3 Categorical colour perception

Categorical colour perception (CCP) is the phenomenon whereby the colour categories in a language influence colour perception. These categorical biases occur in high-level perception, which incorporates attention, memory and evaluation (Witzel, 2019). Therefore, it can be debated whether the effects are strictly perceptual, however high-level perception is defined in a broader sense as direct responses to a present stimulus (sometimes referred to as categorical facilitation to avoid confusion; Witzel & Gegenfurtner, 2015). Although categorical biases may exist in the encoding of colours or in the response, biases are observed using a variety of response methods, suggesting that they are not driven by a decision made during a specific response method, but rather represent how stimuli are interpreted.

To measure CCP, researchers measure reaction times or response accuracy in perceptual discrimination tasks for pairs of colours within and across category boundaries. Participants tend to discriminate between two stimuli that straddle a category boundary (e.g. G2 and B2 in Figure 1.3a) more accurately and faster than two stimuli that belong to the same category (e.g. G1 and G2 in Figure 1.3a), even when the number of JNDs between the pairs is equivalent (Franklin, Drivonikou, Bevis, et al., 2008; Hu et al., 2014; Roberson et al., 2009; Roberson & Davidoff, 2000). As such, performance in discrimination tasks is better when discriminating between colours in different categories than colours in the same category.

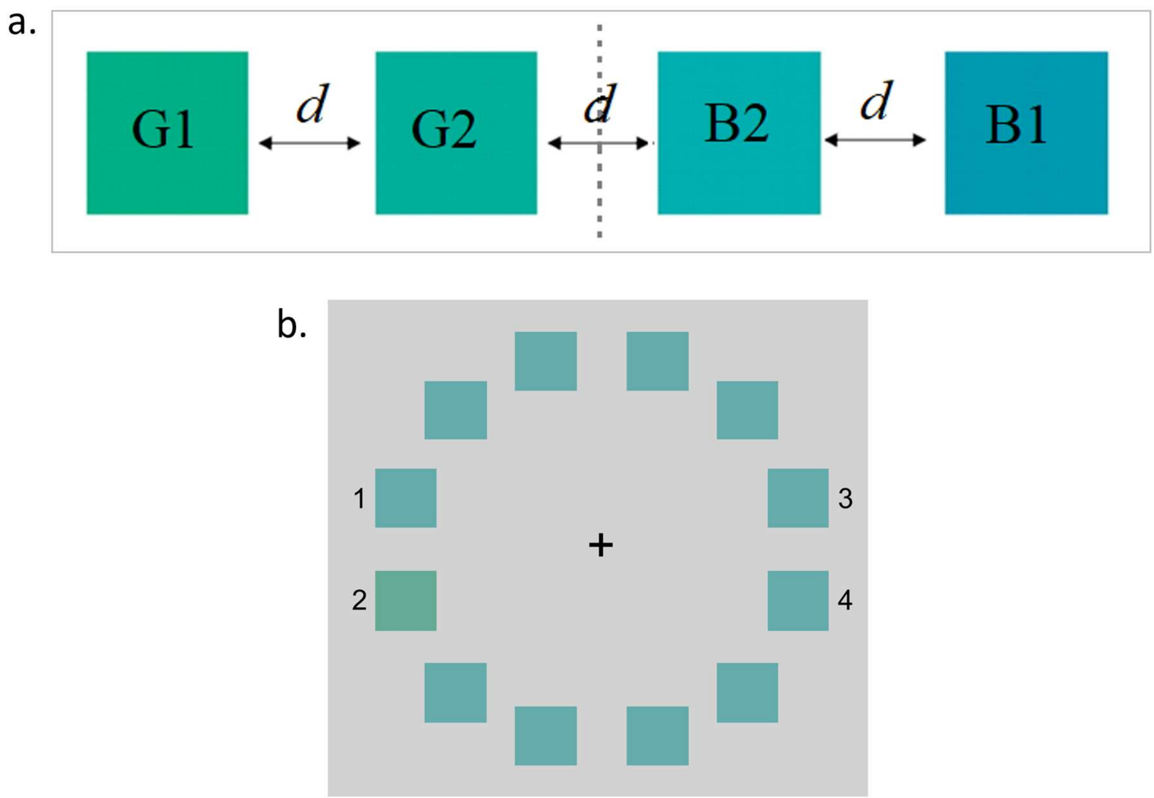


Figure 1.3 *Experimental design for a lateralised CCP task across the blue-green category boundary.(a) d denotes the number of JNDs between the colours, which is equivalent between all four shades. The dashed line in the centre shows the category boundary between green and blue. (b) Sample display for a visual search task, where the target occupied positions 1, 2, 3 or 4. In this example, the target is in position 2, showing a cross-category target in the left visual field. Reprinted from "Newly trained lexical categories produce lateralized categorical perception of color." By K. Zhou, L. Mo, P. Kay, V. P. Kwok, T. N. Ip, and L. H. Tan (2010). Proceedings of the National Academy of Sciences, 107(22), pp. 9974-9978. Copyright 2010 National Academy of Sciences.*

Comparing across languages, CCP is only observed for categories in the speaker's language. Roberson et al. (2000) compared learning and memory for colours between English speakers and Berinmo speakers (a language with 5 BCTs). English has category boundaries that don't exist in Berinmo (such as blue-green), and Berinmo has category boundaries that don't exist in English. Roberson et al. (2000) used a two-alternative forced choice (2AFC) task, where participants were shown a target stimulus, followed by two test stimuli. The task was to identify which of the test stimuli matched the target. Performance accuracy was higher when the test colour straddled a category boundary in the participants' own language, but not for category boundaries that did not exist in their language, demonstrating CCP.

To test whether CCP can be learnt, Özgen and Davies (2002) trained English speaking participants to categorise colours across a new category boundary which did not exist in English, for example by splitting green colours into two novel categories. Participants were trained using a task where they organised colours into the novel categories and were given feedback. Discrimination was tested using a change detection task, where pairs of stimuli are displayed successively with an interstimulus interval of 5s. Participants responded whether the two stimuli were the same or different, i.e. detected a change in the stimulus. After training, discrimination across novel category boundaries was improved, showing that colour perception is malleable and subject to influence by learned colour categories. Since the novel categories were artificially trained, and bore no resemblance to the participants' native language or other natural languages, this CCP effect can be entirely attributed to language learning, rather than a biological determinant.

Zhou et al. (2010) also trained participants to learn new lexical colour categories, and produced a lateralised effect of CCP. A visual search paradigm was used, where participants were shown 12 coloured squares organised in a ring (Figure 1.3b). The target square was a different colour to the 11 distractor squares. Participants responded whether the target was on the right or left visual field of the circle. After training, reaction times were faster in the right visual field for targets across novel category boundaries, demonstrating a lateralised CCP. Lateralised CCP was not observed in the control group, indicating that the effect is entirely a result of learning new categories.

Lateralised CCP refers to the phenomena whereby categorical effects on colour perception occur preferentially for stimuli presented in the right visual field. The right visual field projects to the left hemisphere of the brain, which is dominant for most linguistic tasks (Hellige, 2001). Transfer of visual information across the corpus callosum takes time, and therefore linguistic processing of information in the right visual field is faster than the left visual field. As such, the lateralised performance in speeded colour discrimination tasks to the right visual field suggests that language influences perception (Drivonikou et al., 2007; Franklin, Drivonikou, Bevis, et al., 2008; Gilbert et al., 2006; Roberson et al., 2008; Roberson & Davidoff, 2000; Winawer et al., 2007; Zhou et al., 2010).

Researchers have also found an effect of verbal interference on CCP (Gilbert et al., 2006; Roberson & Davidoff, 2000; Winawer et al., 2007). Participants performed a verbal interference task during the colour discrimination task, such as silently rehearsing an eight digit number which was tested at the end of the block (Gilbert et al., 2006). The verbal interference task increased demands on verbal working memory during colour discrimination and disrupted CCP. Lateralised effects of CCP disappeared, such that performance in the left and right visual field is equivalent. Non-verbal interference tasks had no effect on CCP, suggesting that it is not just an effect of increased cognitive demand (Gilbert et al., 2006; Roberson & Davidoff, 2000; Winawer et al., 2007). This implies that lateralised CCP is a result of verbal processing of colours during a task that does not explicitly require any verbal processing, lending further support that language is involved in perception and discrimination of colours.

The lateralised effect of CCP is observed in both adults and children. However, for pre-linguistic infants, the effect is lateralised in the opposite direction: towards the left visual field and right cerebral hemisphere (Franklin, Drivonikou, Bevis, et al., 2008). As infants learn to identify colour terms and distinguish category boundaries linguistically, CCP switches to dominate in the left hemisphere (Franklin, Drivonikou, Clifford, et al., 2008), suggesting that it is the acquisition of colour terms in language that causes a lateralised effect of CCP in adults.

In all of the previous studies, discrimination was measured across single category boundaries using 2AFC tasks. This is useful to measure whether specific category boundaries have an effect on perception; however the implications are limited, as the perceptual structure has not been measured across multiple categories. To measure the biases in colour perception for a hue circle spanning across blue, green, yellow, orange, pink and purple, Bae et al. (2015) conducted colour naming and colour estimation tasks. In the colour naming tasks, participants were asked to label 180 colours equally spaced across the hue circle, and to identify the best example colour for each label. In the colour estimation tasks, participants were asked to estimate a target colour using the hue circle. Colour estimation was tested with a 900ms delay between target and response, or with no delay (the target was displayed until a response was selected). Using responses from the colour naming task, Bae et al. (2015) identified category foci and boundaries across the hue circle. Response frequency in the colour estimation task was biased towards category foci and away from category boundaries. This categorical effect was observed both with and without a delay, although the categorical effect was greater with a delay. They concluded that there are categorical biases in colour perception, which originate in perception and are amplified by memory demands. This was taken as evidence that perception and memory rely on both continuous and categorical estimates of colour.

Persaud and Hemmer (2014) also conducted a colour naming and colour estimation task across a hue circle. However, in the colour estimation task, participants were shown multiple coloured shapes consecutively, and tested on their colour memory using the stimulus shape as a cue. This increased the memory demands of the colour estimation task compared with Bae et al. (2015). The frequency of responses was biased towards category foci and away from boundaries, further supporting the conclusions made by Bae et al. (2015). Testing only blue and green hues in the hue circle, Cibelli et al. (2016) also produced similar results in a colour estimation task, observing categorical biases in response frequency towards category foci for blue and green, and away from the blue-green category boundary.

A Bayesian model can account for categorical biases observed in colour perception and memory, where prior expectations about colour are used to reduce perceptual uncertainty (Bae et al., 2015; Cibelli et al., 2016; Regier & Xu, 2017). There is a lot of uncertainty in colour perception (Sims et al., 2016). For low-level vision, the dominant noise is due to photoreceptor noise (Vorobyev & Osorio, 1998), whereas for high-level vision colours need to be matched between objects, and therefore the dominant noise is due to failure of colour memory (Baddeley & Attewell, 2009; Bae et al., 2015). Noise increases uncertainty about the sensory signal, and so uncertainty is reduced by combining noisy sensory signals with prior knowledge about the world.

In Bayesian models of CCP, category knowledge is considered as a non-uniform prior over colours, where category foci are more likely and category boundaries are less likely. Under conditions of uncertainty about colour, the prior is integrated with continuous perceptual information, which results in categorical effects on colour perception. When uncertainty is greater, for example increased memory demands or unfamiliar lighting conditions, the prior has a stronger effect on the posterior distribution, thus explaining why larger categorical effects are observed when memory demands are increased for colour estimation tasks (Bae et al., 2015; Cibelli et al., 2016).

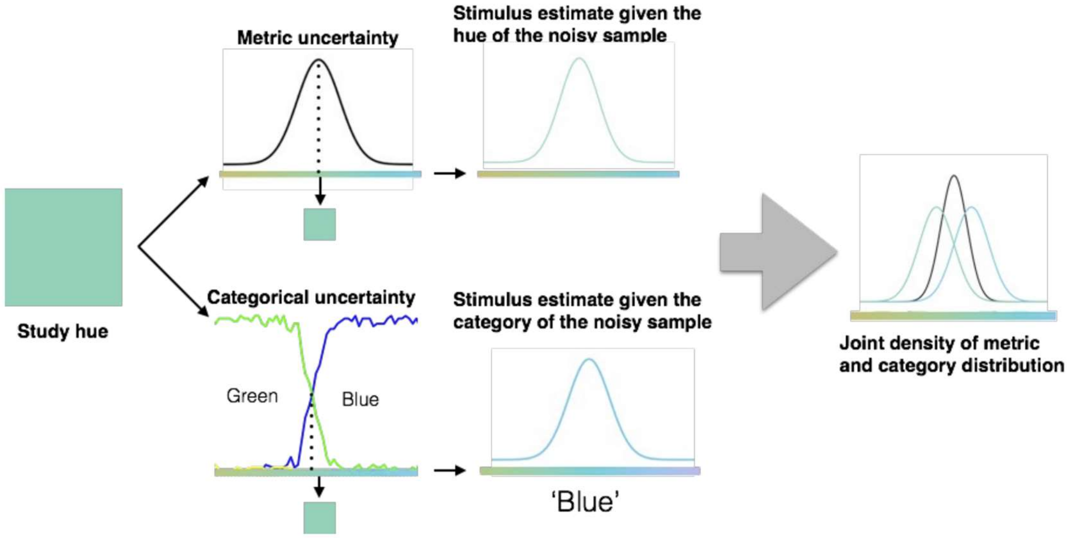


Figure 1.4 *A Bayesian model of colour perception. The study hue is estimated as a joint distribution combining metric uncertainty and categorical uncertainty. Reprinted from "Why some colors appear more memorable than others: A model combining categories and particulars in color working memory." By G. Y. Bae, M. Olkkonen, S. R. Allred, and J. I. Flombaum (2015). Journal of Experimental Psychology: General, 144(4), pp.744. Copyright 2015 APA.*

To test Bayesian models of CCP, responses to the colour estimation task can be modelled as a combination of a stimulus estimate given the noisy sensory signal, and a stimulus estimate given the category assigned to the stimulus (Figure 1.4). The noisy sensory signal is modelled as a Gaussian distribution which centres on the stimulus as a mean, and a variance representing noise in the signal (metric uncertainty). Data from colour naming tasks is used to generate estimates of the category assigned to a stimulus (categorical uncertainty). By combining these estimates, Bae et al. (2015) simulated responses to the colour estimation task. The Bayesian model predicted categorical biases in responses which were significantly correlated with the observed biases in participants' responses. Applying a Bayesian model to different datasets, Cibelli et al. (2016) and Persaud and Hemmer (2014) were also able to simulate categorical biases in responses to colour estimation tasks, thus providing strong support that a non-uniform categorical prior over colours explains the categorical biases observed in colour perception.

Sims et al. (2016) apply rate-distortion theory to this problem, where the visual system represents an optimal solution for minimising cost. Due to constraints on the system's ability to perceive and remember a signal, errors are often made. However, some errors are more costly than others, and therefore Sims et al. (2016) propose that a cost function may be used to direct observers away from the most costly errors and towards the least costly errors. For example, given a hue which is greenish-blue, would misperception of the colour be more costly if biased towards green or blue? Sims et al. (2016) modelled the metric cost (hue angle) and the categorical cost (across category boundaries) of responses to stimuli in Bae et al. (2015). A model which minimised both metric and categorical cost generated a better fit to the observed data than a model which minimised only metric or categorical cost. This effect was mediated by channel capacity, where increased memory demands decrease channel capacity, resulting in larger categorical effects on perception.

It could be argued that the categorical biases observed in colour perception may be due to participants using a verbal labelling strategy to remember the target colour rather than a Bayesian prior. Souza and Skóra (2017) tested this hypothesis by assigning an articulatory suppression task (repeating "bababa" aloud") during a delayed colour estimation task. Categorical biases in responses were observed even during articulatory suppression, where participants relied more heavily on categorical information during increased uncertainty (e.g. larger set size). This suggests that categorical biases are not just a result of verbal labelling, and supports the hypothesis that categorical information serves as a prior to reduce uncertainty in colour perception and memory.

It is clear that perception and memory for colours is not uniform, but rather biased by categorical structure. This effect is mediated by language and memory, where categorical structure is determined by the colour terms in a language, and increased memory demands increase the size of categorical biases on colour perception.

1.4 What underlies universal constraints in colour categorisation?

Evidence for categorical biases in colour perception have been established, but the origin of these colour categories is hotly debated (Regier & Kay, 2009; Siuda-Krzywicka et al., 2019; Witzel, 2019). Traditionally, the argument about the origins of colour categorisation was framed as a debate between universalism and relativism. Universalism is the belief that there are universal constraints on colour categorisation, which result in observed patterns in colour naming across languages (Berlin & Kay, 1991). This position was challenged by relativists, who argued that there is variation in the structure of colour terms across languages, which determines shape of perceptual colour space (Jameson, 2005; Roberson et al., 2000; Whorf, 1956). Therefore it was argued that the structure of a language determines or influences cognition (Whorf, 1956).

As a result of the research that has been conducted over the past 50 years, the universalism vs relativism debate is no longer as dichotomous, since the origins cannot be simply attributed to biological constraints on vision or language constraints. Although language does influence perception, language itself is not a comprehensive explanation for the origins of colour categories, because there are universal constraints on language. It is now generally understood that a complex interaction of constraints and determinants influence colour categorisation and perception. As a result, there are three main themes in the recent literature which attempt to explain universal patterns in colour categorisation. One theme is the universal constraints on language, whereby all languages have the purpose of enabling efficient communication among their speakers. Another theme is the biological constraints placed on colour perception by the human visual system. Finally, there are statistical patterns in the colouration of the environment which dictate the usefulness of colours, and thus may guide perceptual processes and language to enable efficient perception and communication of colours within that environment.

1.4.1 Language constraints

Language is a code which allows an information signal to be transmitted from a signaller to a receiver (Figure 1.5). Transmission can be from one person to another person, or encoded into one’s memory to be remembered at another time. Efficient communication of information is achieved using a signal which has minimal complexity whilst maximising the accuracy (Gibson et al., 2019). Language complexity refers to the level of cognitive demand required to learn and use the language. Accuracy refers to the success with which important information is communicated. Increasing accuracy requires a more complex language, but a simpler language will incur a cost to the accuracy of the signal. This concept of communication efficiency guides the form of linguistic systems, by explaining why language has certain words in semantic domains.

The principles of communication efficiency can be used to investigate universal patterns in colour naming. For a task in which a colour must be communicated from a signaller to a receiver (Figure 1.5), what is the optimal colour naming system to communicate about colours? The complexity of a colour naming system is measured by the number of colour terms in the language, and the accuracy of a colour naming system is measured as the amount of uncertainty about colour which is reduced by the colour terms in a language. In a communication game, the receiver initially has maximum uncertainty about the colour *t*, as an estimate would be random (Figure 1.5). The colour term *w* transmitted by the signaller will provide information about the colour *t,* which will reduce the receiver's uncertainty about *t*. If there are 330 colour chips, then the most accurate code would contain 330 colour terms, and thus a colour term will reduce the receiver's uncertainty by 100%. Although very accurate, such a large lexicon is highly complex and difficult to learn and remember. Therefore, a more efficient lexicon will have an optimal trade-off between complexity of the lexicon and accuracy of the signal.

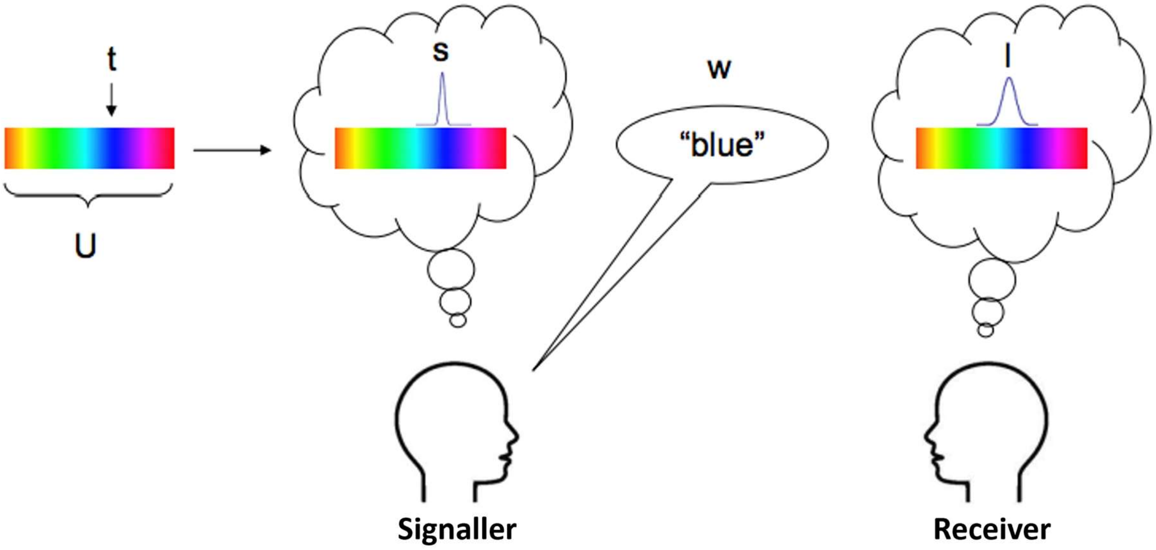


Figure 1.5 *A scenario demonstrating the communication of colours. The signaller observes colour t from set of colours U. The signaller forms a mental representation of t as a probability distribution s over U. The signaller uses a colour term w (blue) to communicate the observed colour to the receiver. The receiver forms a mental representation l over U to approximate t given the signal w. Information about colour has been communicated using language. Reprinted from "Word Meanings across Languages Support Efficient Communication." By T. Regier, C. Kemp, and P. Kay (2015). The handbook of language emergence, 87, p.237.*

The colour naming tasks used in the WCS replicate the communication game, as participants named colour chips (signaller) and estimated a colour chip for a given colour term (receiver). As such, naming data in the WCS can be used to assess the communication efficiency of languages.

Regier et al. (2015) compared the communication efficiency of colour lexicons in the WCS with a set of hypothetical colour lexicons. The researchers randomly assigned *n* categories to the 330 Munsell colour chips used in the WCS, and iteratively changed category labels on individual chips to increase the communication efficiency of a hypothetical colour lexicon with *n* categories. This was repeated 20 times from random starting configurations for colour lexicons with *n*=3-11. For each level of language complexity, the accuracy of the colour lexicon was close to the hypothetical optima, suggesting that languages universally are close to maximally efficient for communicating about colours. In addition, the theoretically optimal configurations with 3-6 colour terms followed stages 2-5 in the evolutionary sequence of BCTs (Figure 1.1). Thus, Regier et al. (2015) found evidence that constraints on communication efficiency can explain universal patterns observed in colour naming.

Although some of the colour naming systems in the WCS did closely match the optimal configuration for communication efficiency, this was not true for all of the languages (Regier et al., 2015). A possible explanation for this is that the theoretical configurations were initiated using random configurations for each complexity level, and therefore the optimal accuracy of a language was determined by the number of BCTs. However, languages do not evolve from random configurations for a specific complexity, but rather from low complexity languages to high complexity. Therefore, many of the colour categories observed in languages in the WCS may represent the most informative configuration on a transition between stages of complexity.

Zaslavsky et al. (2018) resolved this issue by measuring language complexity on a continuous scale rather than categorising languages by complexity using the number of BCTs. Complexity was measured as the amount of information provided by colour terms about colour chips, where the more information provided by colour terms, the higher the complexity of the language. Using this new measure, all languages in the WCS were near optimally efficient for a trade-off where maximising accuracy is slightly more important than minimising complexity. Simulated colour lexicons also predicted low consensus colour terms which emerge in languages during the transition towards higher complexity, and therefore provided a much better fit to the colour lexicons in the WCS than the original model in Regier et al. (2015). In addition, Chaabouni et al. (2021) found that neural networks trained to play a communication game (as in Figure 1.5) developed colour lexicons with a similar trade-off between accuracy and complexity as that observed by Zaslavsky et al. (2018) in the WCS. This further supports the argument that the trade-off between accuracy and complexity observed in colour naming systems is optimal to communicate about colours.

Information theoretic methods find good evidence for the principles of communication efficiency guiding colour naming, but these approaches neglect to consider the context in which communication occurs. Models were trained to optimise communication of coloured squares varying uniformly across colour space, but the colours that humans communicate about vary systematically in perceptual structure and in the distribution of colours in the environment. This systematic variation has an effect on the behavioural relevance of colours, for which communication should also be optimised.

1.4.2 Biological constraints

It was traditionally argued that biological constraints which are hard-wired into the human visual system are responsible for cross-cultural similarities observed in colour categorisation (Kay & McDaniel, 1978). Although there is strong evidence for the effects of language on colour categorisation, CCP has also been observed in the absence of language, for example in pre-linguistic infants, patients with brain damage and non-human animals (Siuda-Krzywicka et al., 2019). This research suggests that there may be a biological basis to colour perception that is unrelated to language.

Colour perception is measured in infants using a novelty preference paradigm, where infants are habituated to a target hue through repeated presentation (Franklin & Davies, 2004). Infants prefer novel stimuli and spend more time looking at them, therefore habituation with the stimulus is achieved when looking time is reduced. Following habituation, the infant is presented with a novel test hue, and an increased looking time demonstrates a novelty preference, indicating that the infant has recognised and distinguished the test hue from the target hue. Infants aged 4-6 months show a novelty preference discriminating green from blue, blue from purple and pink from red (Franklin & Davies, 2004). This novelty preference is independent of chromatic discrimination, as infants were able to perceptually discriminate between hues within a category, suggesting that infants were demonstrating CCP in their novelty preference. Franklin et al. (2005) also tested categorical colour perception in pre-schoolers aged 2-4 years, which is the age range for the onset of colour term knowledge. CCP was observed regardless of the infants' colour term knowledge, suggesting that CCP was not related to colour naming.

Skelton et al. (2017) used the novelty preference paradigm to test infants with a mean age of 21.3 weeks. Despite being extremely young, infants still displayed a novelty preference for colours which straddled a category boundary, although the location of the category boundaries were different to those of English speaking adults. The novelty responses were plotted in MacLeod Boynton colour space (and in DKL-space in Siuda-Krzywicka et al., 2019), and four of the category boundaries identified by infants were separated by cardinal axes which correspond to second stage opponent mechanisms in colour vision. This evidence suggests that colour categories do partly originate from the underlying mechanisms in the early visual system that process colour, and could explain universal patterns in colour categorisation.

The issue with the novelty preference paradigm is that inferences are made from the absence of a behavioural change, i.e. it is assumed that infants don't recognise a category boundary when looking time does not increase. However, there are multiple reasons why an infant may not demonstrate a novelty preference for the test hue: they may have forgotten the target hue or simply have been distracted. As a result, the novelty preference paradigm is a less reliable measure of CCP in infants compared to colour matching and discrimination tasks used to test CCP in adults.

In a lateralised target detection task, both adults and infants aged 4-6 months old were faster at detecting a chromatic hue displayed on a chromatic background when the target and background belonged to different colour categories (Franklin, Drivonikou, Bevis, et al., 2008). This study used the same task to measure CCP in adults and infants, so a direct comparison can be made between their responses. CCP was lateralised to the left hemisphere in adults, but to the right hemisphere in infants. As infants acquire colour term knowledge, lateralised CCP switched to the left hemisphere (Franklin, Drivonikou, Clifford, et al., 2008). Given that the left hemisphere is dominant for language processing, these results suggest that there are biological constraints on CCP in infants, which may interact with language as colour terms are acquired.

Categorical colour encoding in pre-linguistic infants has also been observed in functional brain activation (Yang et al., 2016). Using near infra-red spectroscopy whilst presenting alternating green and blue shapes to infants, Yang et al. (2016) found significant brain activation compared to the baseline in response to different-category (G1-B1, G2-B2) trials, but not for same-category (G1-G2, B1-B2) trials. The study therefore confirmed the presence of categorical encoding in coarse regions of the infant brain.

An independence between colour categorisation and language can also be investigated using patients with brain damage. Colour anomia refers to deficits in colour naming as a result of brain damage, despite intact colour perception and intact general language processing (Oxbury et al., 1969). Patients are unable to name colours presented visually, but they are able to sort visually presented colours into colour categories associated with linguistic terms (Fukuzawa et al., 1988; Haslam et al., 2007; Roberson et al., 1999). This suggests that colour categorisation can occur independently of colour naming and language processing.

In addition, it is possible to train animals to categorise colours using operant conditioning. When distinctions between colours are behaviourally relevant to obtain food, chicks and goldfish were able to learn colour categories (Jones et al., 2001; Poralla & Neumeyer, 2006). More recently, researchers have investigated colour categorisation for colour signals which are naturally behaviourally relevant to animals. For example, male zebra finches have beaks which vary in colouration from light orange to dark red, which is a reliable signal of their immunity level (Caves et al., 2018). It is therefore an adaptive signal for female zebra finches to identify healthy males to mate with. Caves et al. (2018) tested colour categorisation along orange and red hues in female zebra finches using Munsell colour chips, and identified categorical patterns in colour perception. CCP has also been observed in blackbirds and robins, who rejected eggs in their nest based on a category boundary which signalled the brown colour of avian brood parasites (Hanley et al., 2017). Altogether, evidence for CCP in non-human animals demonstrates that colour categorisation is possible without language, therefore language alone cannot explain universal patterns in colour categorisation.

Given that CCP is observed independently of language, it is possible that colour categories may originate from the underlying mechanisms in the early visual system that process colour, and could explain universal patterns in colour categorisation. Kay and McDaniel (1978) highlighted that white, black, red, green, yellow and blue are the first six colour terms to be acquired by languages in stages 1-5 of Berlin and Kay's evolutionary sequence, and they are also the opponent colour pairs which occur during the second-stage processing of photoreceptor signals in the LGN. Kay and McDaniel (1978) suggested that early stage BCTs represent the opponent colours, and later stage BCTs are a mixture of these opponent colours. As a result, the literature focused on using colour opponency theory to explain the origins of universal constraints in colour naming (Jameson & D'Andrade, 1997; Saunders & van Brakel, 1997).

However, investigations into the relationship between category boundaries and mechanisms in the early visual system in adults have failed to find strong support for this hypothesis (Emery et al., 2017; Malkoc et al., 2005; Witzel & Gegenfurtner, 2013, 2018; Wuerger et al., 2005). In these studies, the peak sensitivities of cone opponent channels are compared with colour naming data. Some studies do find that the cardinal cone-opponent axes delineate the boundary between green and blue (Malkoc et al., 2005; Witzel & Gegenfurtner, 2013, 2018), however there is no evidence to suggest that opponent mechanisms can explain the boundaries across red, green, yellow and blue. Therefore colour categorisation cannot be entirely explained by the neurophysiology of the visual system.

It is possible that a combination of biological and social constraints contribute to universal patterns in colour categorisation. Baronchelli et al. (2010) generated synthetic languages and simulated independent groups of interacting individuals. Cultural transmission in groups who were endowed with human JNDs developed universal patterns in colour categorisation similar to the colour lexicons observed in the WCS. Neutral groups who were not endowed with JNDs (therefore had a uniform colour percept) developed colour lexicons similar to randomised data in Kay and Regier (2003). These simulations, in addition to Skelton et al. (2017) identifying that infant colour categorisation coincides with opponent mechanisms, suggest that biological constraints on the visual system may contribute to the emergence of universal patterns in colour categorisation, but only when combined with constraints associated with language. The difficulty in disentangling biological and linguistic constraints may explain the lack of evidence for a link between mechanisms in the visual system and universal patterns in colour categorisation. Additionally, while language is not vital for colour categorisation to be observed, it is an important determinant of the categories observed across languages.

1.4.3 Ecological constraints

Methods which investigate linguistic or biological constraints tend to measure colour categorisation for colour patches, without considering the context in which colours are categorised. In the real world, the colours that humans perceive and communicate about are strongly associated with objects (Palmer & Schloss, 2010). As such, the distribution of colours across objects in the environment (visual diet) may place an ecological constraint on perception, language and colour categorisation.

The colours of objects are behaviourally relevant to animals for foraging, reproduction and avoiding predators (Cuthill et al., 2017), for example the colour of a mango indicates its ripeness, or the colour of a potential mate's complexion is an indicator of their health (Whitehead et al., 2012). The objects which are behaviourally relevant, and the colours associated with them, will depend on the environment and the animal's visual system. For example, in the arctic the colour white is much more prevalent, and therefore it is more useful to discriminate between different shades of white (Fortescue, 2016). In addition, human trichromatic vision provides an advantage for detecting behaviourally relevant fruits from leaves (Regan et al., 2001). As such, vision and language are tightly related to the behavioural relevance of object colours.

If there are systematic patterns in colouration across behaviourally relevant objects, then colour categories may be formed in a high-level structure to guide perceptual processes towards visual stimuli which are the most important. Rate-distortion theory proposes that some errors in colour perception are most costly than others, and this cost function determines categorical boundaries by directing observers towards the least costly errors in perception (Sims et al., 2016). When using colour signals for object recognition, for example selecting a ripe fruit, it is beneficial to perceive colour categorically, so that the fruit can be easily recognised as unripe, ripe or overripe, thus reducing the likelihood of misrecognising the ripeness of a fruit and any costs associated with doing so. As such, colour categorisation may serve the function of reducing the likelihood of costly misrecognition, and universal patterns in colour categorisation may serve as the optimal solution for this problem.

Ecological constraints on colour categorisation can also explain instances where colour categorisation is observed without language. Animals do learn associations between behaviourally relevant objects and their colour patterns to improve recognition of rewards or threats (Caves et al., 2018; Hanley et al., 2017). Pre-linguistic infants may also have acquired colour categories by learning associations between colours and objects. Younger (1990) demonstrated that pre-linguistic infants are able to detect and learn correlations among object attributes, e.g. toy animals with furry tails also have antlers. These learnt correlations are the foundation of category formation, and suggest that infants may also be able to identify statistical tendencies of object colours and thus form colour categories without language. In addition, Kimura et al. (2010) found that 6-8 month old infants showed a preference for images of faces and fruit which were correctly coloured versus inappropriately coloured (e.g. blue banana), whereas 5 month olds showed no preference. This suggests that relationships between colours and objects are learnt around 6 months old, before language emerges.

The behavioural relevance of features in the environment have also been shown to influence visual working memory. In a change detection task, participants' memory performance improved based on the probabilistic context in which the stimuli were sampled, such that performance was better for changes that were most common (Bates et al., 2019). In addition, participants learnt to group the stimuli based on the similarity of their features, and performance was better for changes that crossed the novel category boundary (Bates et al., 2019). Therefore, humans are receptive to the statistics of features in the environment, and use these statistics to guide cognition. Similarly, humans may learn the statistics of colours in the environment to guide cognition towards the most probable stimuli, thus resulting in categorical perception of colours.

If colour categories are acquired from the distribution of colours in the environment, then the BCTs in a language may vary with the behavioural relevance of colours in different environments or cultures. However, if the behavioural relevance of colours is shared across environments and cultures, then similar patterns in colour categorisation would be predicted.

To test whether there is a relationship between language and the environment, Baddeley and Attewell (2009) calculated the optimal number of lightness terms required to communicate reflectance information, by achieving a trade-off between language complexity and accuracy. Surface reflectances were measured from four different types of visual environment: deciduous woodland, beach, urban and domestic interior. Using a communication game (as in Figure 1.5), a signaller communicated reflectances to a receiver. Three lightness terms, corresponding to white, grey and black, were the optimal number to efficiently communicate information about reflectances in the environment. Therefore, BCTs for variation in lightness were explained by modelling colour categorisation as a product of constraints from communication, the environment and perceptual uncertainty. Applying similar methods to investigate the origins of all 11 BCTs requires many more samples from the environment, in order to obtain a good estimate of the distribution of colour across hue, saturation and brightness.

Yendrikhovskij (2001) collected the statistics of 630 natural images in CIELUV colour space, and identified a non-uniform distribution: low saturation colours were more frequent than high saturation colours. A clustering algorithm was applied to the colours which minimised similarity within a cluster and maximised similarity between clusters. A configuration of the cluster centres for natural images was more similar to the foci of BCTs than a configuration of cluster centres for a uniform distribution of colours, suggesting that universal colour categories do represent the optimal solution for categorising colours in the environment.

Belpaeme and Bleys (2009) replicated this study using a different dataset with images of natural and urban environments, as well as a control environment where pixels were selected randomly from RGB space. All pixels were then converted into CIELUV and CIELAB for testing. Low saturation colours were again more common in pixels sampled from images, whereas frequency of colours was fairly uniform for the control environment. A clustering algorithm identified colour categories in the distribution of pixel colours, and positive correlations were observed between clusters for natural, urban and control environments. Since clusters related to BCTs were also observed in the control environment, this suggests that the clusters are not entirely dependent on the statistical distribution of colours in our environment, but also due to biases in the colour appearance model and the clustering algorithm.

However, Yendrikhovskij (2001) and Belpaeme and Bleys (2009) used the frequency of pixel colours to estimate environmental colour statistics, operationalising that the usefulness of colours are determined by colour frequency. The colours that humans perceive and communicate about are strongly associated with objects (Palmer & Schloss, 2010), and so colours also need to be analysed within a context where they are useful for communicating about objects in the environment.

Gibson et al. (2017) argued that objects are more useful to humans than backgrounds, therefore there may be a greater need to communicate about object colours than background colours. To investigate this hypothesis, human observers segmented objects from their background for over 20,000 photographs of salient objects. The researchers then measured the difference between the colour statistics of objects and backgrounds. For both naturally and artificially coloured objects, objects tended to be warm coloured (red/yellow/orange) and backgrounds were cool coloured (blue/green). Gibson et al. (2017) then looked at colour naming data for the 110 languages in the WCS, along with original data from 3 groups (Tsimane, Bolivian-Spanish and English), and found that across all languages, colour terms supported more precise communication of warm colours compared to cool. Essentially, languages tend to have more terms to describe warm colours than cool colours. Gibson et al. (2017) concluded that the universal warm-cool asymmetry of colour terms reflects a greater need to communicate about warm colours, as warm-coloured objects are more behaviourally relevant than cool-coloured backgrounds.

Rosenthal et al. (2018) conducted a follow-up study, where a classifier was trained to distinguish whether pixels belonged to objects or backgrounds. For each image in the dataset used in Gibson et al. (2017), the mean pixel value was calculated across the object and the background. The classifier successfully discriminated between objects and backgrounds, and between animate and inanimate objects, using mean pixel colour. This suggests that there are statistical patterns in environment colouration, which may be exploited by colour categorisation to facilitate object detection and recognition.

Zaslavsky et al. (2019) argued that the warm-cool asymmetry in colour naming is represented by the structure of perceptual colour space and the colour statistics of the environment. To test this, they converted Munsell colour chips into CIELAB space, and applied *k*-means clustering to create hypothetical colour naming systems for each language in the WCS. Zaslavsky et al. (2019) noted that these hypothetical colour naming systems determined by perceptual colour space also exhibit a warm-cool asymmetry, with a bias towards more precise communication for warm colours. As such, perceptual structure may also explain the warm-cool asymmetry observed by Gibson et al. (2017) in colour lexicons universally. However, it is still possible that these asymmetries originate from the environment, as perceptual systems may have adapted to the frequency with which particular colours appear in the world, and languages to the frequency with which they must be communicated. In addition, the frequency with which colours appear and the frequency with which they must be communicated are likely related; for example, colours which occur less frequently may be a more important signal for communication.

In these previous studies it was assumed that objects are more behaviourally relevant than backgrounds, and this is the factor which determines the important signals for communication across languages. However, these signals are likely to vary across languages, and therefore Twomey et al. (2020) estimated the important colours for communication for individual languages, without making any assumptions about what determines an important signal. Using mappings from colour terms to colour chips (w to l in Figure 1.5) in the WCS, they modelled the colours which are communicated most precisely within a language, and thus inferred the colours which are most important for communication. The model was used to predict mappings from colour chips to colour terms. Significant cross-cultural variation was observed in communication across the colour space. This variation was related to geographic location and local biogeography of the culture, suggesting that the colour lexicon is adapted to the colour statistics of the environment.

In addition, when averaging communication efficiency of colours across languages, warm colours were communicated more efficiently than cool colours (Twomey et al., 2020), thus supporting previous research suggesting that a warm-cool asymmetry in colour lexicons supports more efficient communication of behaviourally important warm-coloured objects (Gibson et al., 2017; Rosenthal et al., 2018; Zaslavsky et al., 2019). The colours of unripe, mid-ripe and ripe fruit also occupied regions of the Munsell colour space corresponding to low, medium and high efficiency of communication respectively (Twomey et al., 2020). This suggests that colour lexicons support more efficient communication about behaviourally relevant ripe fruit, further supporting the theory that colour lexicons are adapted to the environment to improve perception of behaviourally relevant stimuli.

However, variations in lightness interacted with the warm-cool asymmetry, where the asymmetry was mostly observed for colours with a medium lightness, and pastel green/blue were communicated less efficiently than dark green/blue (Twomey et al., 2020). In addition, Twomey et al. (2020) did find a small but significant correlation between Munsell chroma and the average distribution of communication efficiency across colours. This highlights the importance of sampling stimuli from a perceptually uniform colour space, and controlling all three dimensions of colours.

1.5 Summary and Conclusion

Colours vary on a continuum across hue, saturation and lightness. In order to communicate about colours which are behaviourally relevant, the continuous colour space is categorised. Colour categorisation is not structured at random, but rather the structure serves an adaptive purpose to improve object detection, object recognition and communication. As a result, there are universal patterns in colour naming and colour categorisation.

Explaining the universal constraints on colour categorisation is not trivial, as cross-cultural patterns in colour naming are closely related to the structure of perceptual colour space, and the statistics of colours in the environment. While some patterns in environmental colour statistics relating to the warm-cool asymmetry in colour naming and perceptual space have been identified, there is currently no clear evidence for a relationship between the colour statistics of the environment and the categorical structure across the full set of BCTs. Part of the difficulty in this problem is in generating a model which simulates the adaptive functions served by colour categorisation. A unified approach has focused on the use of colour for object detection using classification of objects vs backgrounds. Colour is useful for segmenting objects from backgrounds in low-level colour perception, but for high-level colour perception, colour is also useful for recognising specific objects. The origins of BCTs may be identified by studying the information provided by colours in high-level perception. This is an approach that is yet to be explored.

To explore this, two tasks are used in this project: a colour estimation task to measure the geometry of high-level colour perception, and an object recognition task to measure the distribution of colour across objects. In Chapter 2, a replication of Bae et al. (2015) is conducted to test whether there are categorical biases in colour perception across hues. In Chapter 3, this experimental design is extended to measure colour estimation across hue and saturation. Participant's responses are used to identify the regions of 2D chromaticity space that attract and repel responses, thus formalising the geometry of high-level colour perception.

In the next three chapters, colours are sampled from images of objects to represent the colour statistics in the environment, and a model is trained to recognise objects using the sampled colour. In Chapter 4, kernel density estimation is used to estimate probability distributions of colours given object classifications in ImageNet. In Chapter 5, a deep neural network is trained to recognise objects in ImageNet for a given colour. In Chapter 6, the same deep neural network is trained to recognise objects for an alternative image dataset (Open Images V6) to test whether the patterns observed in ImageNet data were replicated. The structure of information provided by colour about objects is compared to the geometry of perception colour space to identify whether the colour statistics of objects in the environment form the basis of a non-uniform prior in high-level colour perception. The implications of these results are discussed in the final chapter.

2 A replication study: why some colours appear more memorable than others (Bae, Olkkonen, Allred, & Flombaum, 2015)

There is a considerable amount of uncertainty during colour perception and memory. Previous research has made assumptions that uncertainty is uniform across the continuous colour space, but there is evidence to suggest that categorical biases exist in colour perception and memory. To test the uniformity of uncertainty across continuous colours, Bae et al. (2015) measured response frequency in a colour estimation task where participants were required to estimate a target colour using a hue circle. Responses were biased towards category foci and away from category boundaries. In this chapter, a replication of this colour estimation task is reported. The original findings are successfully replicated: response frequencies were not uniform, but rather biased by category structure. In addition, a novel finding involving a systematic error in estimating green targets as blue was observed. The results provide further evidence for categorical biases in colour representation.

2.1 Introduction

Visual input is ambiguous due to the complexity of viewing conditions and noisy neural processing. In addition, memory capacity reduces the accuracy with which information is stored and reproduced (Sims, 2015). This results in a considerable amount of uncertainty during colour perception and memory. As such, a task as simple as identifying a ripe banana can be computationally difficult, even though the action feels effortless. To reduce uncertainty, an optimal observer will use prior knowledge to make unconscious inferences about ambiguous visual input (Helmholtz, 1867). These priors bias perception towards the most likely outcomes and away from those that are unlikely. For example, a prior expectation that bananas are yellow biases appearance of bananas towards yellow (Adeyefa-Olasupo & Flombaum, 2018).

Research investigating the mechanisms behind colour perception and memory commonly use colour estimation paradigms, where participants are required to reproduce a target colour after presentation (Fougnie & Alvarez, 2011; Wilken & Ma, 2004; Zhang & Luck, 2008, 2011). The difference between target colour and response colour is measured to determine the level of uncertainty in these mechanisms. This task reflects similar challenges in everyday life, where information about colour is encoded into memory to be remembered later for object detection or recognition.

Variables in the colour estimation task are modified which influence uncertainty, such as delay length between presentation and response or the number of items in each trial. Representations of colour memory were modelled as noisy, continuous estimates of colour, where responses are normally distributed around the target colour, and the variance of the distribution is a measure of uncertainty (Figure 2.1; Zhang & Luck, 2008). Therefore, it was assumed that the variability of responses is roughly similar across all colours sampled, so responses in each condition were averaged across different colour stimuli.

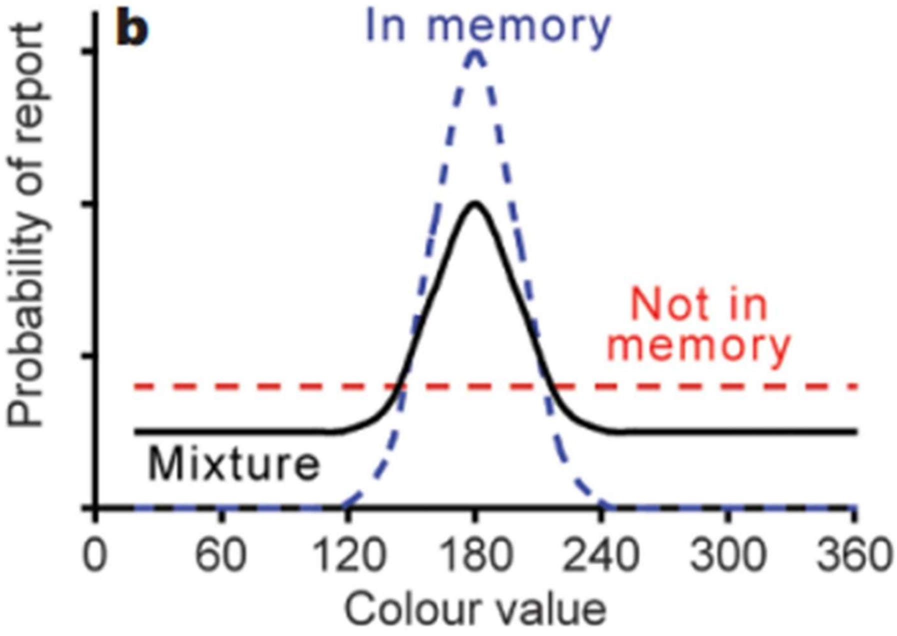


Figure 2.1 *Continuous model of colour memory for a sample colour at 180°. Performance in a colour recall task is represented by a mixture of a continuous estimate of the colour (blue) and uniform uncertainty (red). Reprinted from "Discrete fixed-resolution representations in visual working memory." By W. Zhang and S. Luck, 2008, Nature, 453(7192), p.233-235. Copyright 2008 by Nature Publishing Group.*

However, it is not correct to assume that uncertainty is uniform across colour stimuli. Although colours are continuous, humans categorise the continuous colour space in order to communicate about the colours in the environment, and the colour categories in a language influence colour perception (Hu et al., 2014; Özgen & Davies, 2002; Roberson et al., 2009; Roberson & Davidoff, 2000). This suggests that participants representation of colour is not uniform across the colour space evaluated, but rather biased by categorical structure. Therefore, uncertainty about colour may vary across stimuli in a colour estimation task.

In colour discrimination studies, participants' responses were collected using a forced-choice paradigm, which involves partitioning the continuous colour spectrum into discrete steps (Hu et al., 2014; Özgen & Davies, 2002; Roberson et al., 2009; Roberson & Davidoff, 2000). This methodology may encourage participants to rely on discrete, categorical encodings of colour during their response, thus producing categorical biases in responses. In addition, differences were examined for responses across a single category boundary, such as the blue-green boundary. By sampling a very small range of the colour space, the results cannot be generalised to a categorical representation of colours across the full range of hues.

To tackle these limitations and test previous assumptions about the uniformity of colour in visual working memory, Bae et al. (2015) conducted colour naming and colour estimation tasks using the full hue circle and examined colour specific differences in participants' responses. They tested three hypotheses: 1) colour specific biases exist in working memory; 2) these biases originate in perception; and 3) observers use both categorical and continuous information in colour estimation. Stimuli were 180 colours sampled from a hue circle varying in CIELAB space ($L^*= 70$, $a^*= 0$, $b^*= 0$, radius of 38), which span the colour categories; blue, purple, pink, orange, yellow and green. These colours varied only in hue, whilst saturation and luminance remained constant.

In the colour naming tasks, participants were asked to categorise colours by labelling colour patches sampled from the hue circle (Figure 2.2a) and then identified focal colours on the circle for the six colour categories (Figure 2.2b). From this, Bae et al. (2015) identified category boundaries and category foci in the hue circle. In the colour estimation tasks, participants were presented with a colour patch and asked to reproduce the colour from the hue circle, both with and without a 900ms delay (Figure 2.2c). The undelayed task measures whether biases are rooted in perceptual processes, and the delayed task measures the influence of working memory on the task.

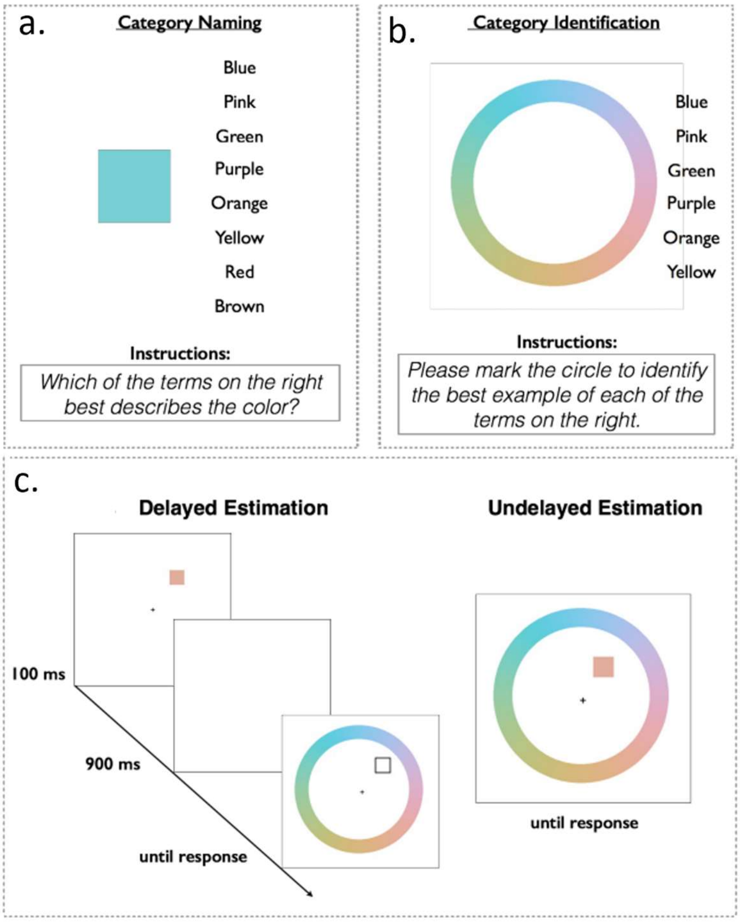


Figure 2.2 *Procedure for the four experiments conducted by Bae et al. (2015). Reprinted from "Why some colours appear more memorable than others." By G.-Y. Bae, M. Olkkonen, S. Allred, and J. Flombaum, 2015, Journal of Experimental Psychology: General, 144(4), p.744-763. Copyright 2015 by APA.*

For both the delayed and undelayed colour estimation tasks, Bae et al. (2015) observed colour-specific biases in response frequency towards the category foci and away from the category boundaries (Figure 2.3a). As the biases occur without a delay, and are correlated with response frequencies in delayed responses, this supports the hypothesis that the biases originate in perception. The evidence therefore suggests that perceptual colour space is not uniformly distributed, but rather biased by category structure. Reproducing a perceived and remembered hue relies on both continuous and categorical information, thus producing stimulus specific biases in colour estimation.

The current experiment aimed to test if these findings by Bae et al. (2015) were replicable. The methods concentrate on the delayed task as the effects are similar to the undelayed task but the effect sizes are larger and thus easier to measure. Stimuli were sampled from a 130° portion of the 360° hue circle to ensure that enough observations were collected per stimulus to measure the effect. The colour samples spanned across the orange, yellow and green categories.

2.2 Method

2.2.1 Participants

Fifteen students aged between 23 and 56 were recruited from the University of Bristol, of which 8 were female and 7 were male. All participants reported normal or corrected to normal vision, with no colour vision deficiencies. The experiment was approved by the School of Psychological Science Research Ethics Committee.

2.2.2 Equipment

The experiment was conducted in a dark, quiet room where the monitor was the only light source. Stimuli were presented on a ViewPixx/3D CRT monitor with a spatial resolution of 1920 x 1080 pixels, subtended 47.42° by 27.75° of visual angle. The monitor was calibrated according to standard methods (Hansen & Gegenfurtner, 2013) using a Minolta CS-100 photometer. Luminance and chromaticity of achromatic, red, green and blue channels were measured for 8 calibration points between minimum and maximum intensity. The gamma value was calculated in the Monitor Centre in Psychopy, which linearised the intensity of the colour guns. The display had an average luminance of 47.5cd/m^2. Participants sat at a viewing distance of 60cm with a fixed chin rest. A standard computer mouse was used as an input device.

2.2.3 Stimuli

Twenty seven equally spaced stimuli between 50° and 180° were chosen, varying in CIELAB space ($L^*= 70$, $a^*= 0$, $b^*= 0$, radius of 38). This ring is based upon that used by Bae et al. (2015) in their delayed estimation task, although only a 130° portion of the ring was tested. The colour samples spanned across the orange, yellow and green categories, as identified in the category identification task by Bae et al. (2015). RGB values were calculated from the CIELAB coordinates using the colormath module implemented in Python. Stimuli were presented on a uniform grey background, which was the centre point of RGB colour space.

2.2.4 Procedure

Each trial began with a white fixation cross (1.57° x 1.57°) displayed in the centre of the screen for 500ms. Following this, the study stimulus (2.62° x 2.62° coloured square) appeared in one of eight locations (5.23° from fixation) for 100ms. After a 900ms delay, during which the screen remained blank, the hue circle (12.93° radius and 3.82° thick) was presented in the area surrounding the stimuli, with a fixation cross and a square patch in the same location as the previous study stimulus. On each trial, the hue circle was randomly rotated to prevent position-colour associations. The participant used the mouse to browse the hue circle, which simultaneously changed the colour of the square patch. Participants clicked on the desired colour to submit a response, which was followed by a 500ms inter-trial interval.

Each block consisted of 54 trials, during which each stimulus was presented twice, in a random order. After five practice trials, participants completed eight blocks of trials, which took around 40minutes to complete. This produced 16 measurements per colour per participant, and 240 observations per stimulus overall.

2.3 Results

The hue angle of participants' responses was measured, and response frequencies across the hue circle were plotted on a histogram (Figure 2.3b). Each stimulus was presented 240 times, but the frequency with which participants responded was not uniform. Responses were biased by category structure, with clusters around an orange-yellow peak at 90°, and a green peak at 150°. There was a category boundary between yellow and green at 115° as evidenced by a local minimum in response frequency between the two peaks. Additionally, among the noise across untested hues, there was a peak in responses at 260° around the blue category, despite not presenting any blue hues to participants.

To further investigate the difference between target and response hues, the responses were fit with a mixture of three von Mises distributions for each individual stimulus. The von Mises distribution is a circular equivalent of the Gaussian probability distribution, where probability density is plotted around the circumference of a circle. A mixture of three von Mises distributions provides flexibility for the model to fit multimodal data. Histograms of the data revealed multiple peaks in probability density around the hue circle (Figure 2.4), therefore using a mixture model was appropriate to accurately fit multiple peaks and noise within the data.

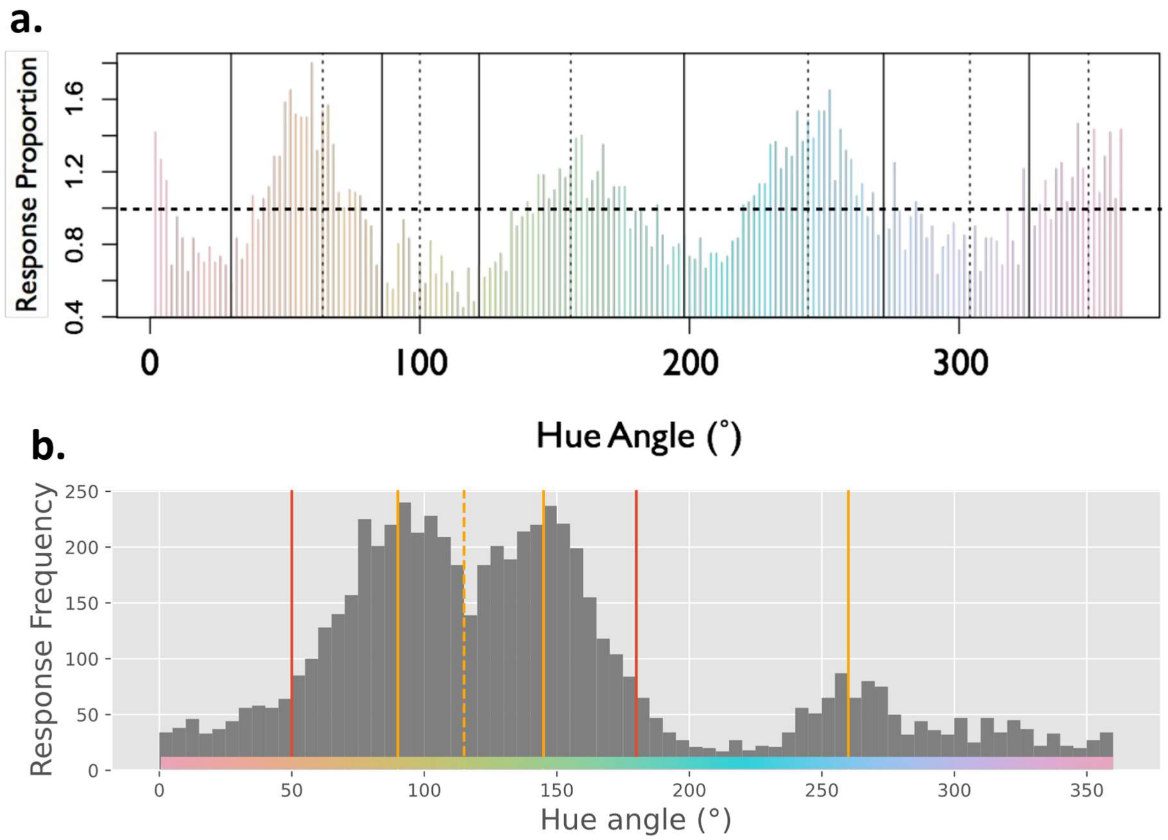


__Figure 2.3__ Response frequencies for each hue in the delayed colour estimation task in (a) Bae et al. (2015), and (b) the current replication. In (a), the vertical dashed lines indicate category foci, and the vertical solid lines indicate category boundaries, as identified in the colour naming studies in Bae et al. (2015). In (b), only the colours within the two vertical red lines, from 50° to 180°, were presented as stimuli. The vertical orange lines indicate peaks in response frequency, and the vertical orange dashed line indicates a trough. Adapted from "Why some colours appear more memorable than others." By G. Bae, M. Olkkonen, S. Allred, and J. Flombaum, 2015, Journal of Experimental Psychology: General, 144(4), p.744-763. Copyright 2015 by APA.

Graphs a-c in Figure 2.4 demonstrate how the von Mises mixture model fit to three of the target stimuli. As a mixture of three distributions was used, the model predicted three mean responses for each target hue. Target hue angle was plotted against response hue angle for each mean of the von Mises model. The first mean closely predicts the target hue angle, demonstrating that participants responded accurately to the colour estimation task (Figure 2.4d). The plot for the second mean shows a systematic error in responses to green stimuli, with participants responding blue, despite no blue stimuli being present in the sample (Figure 2.4e). Responses to yellows and oranges are also biased towards orange and pink. This dichotomy of responses being biased in opposite directions produces the category boundary in response frequencies at 115° between yellow and green.

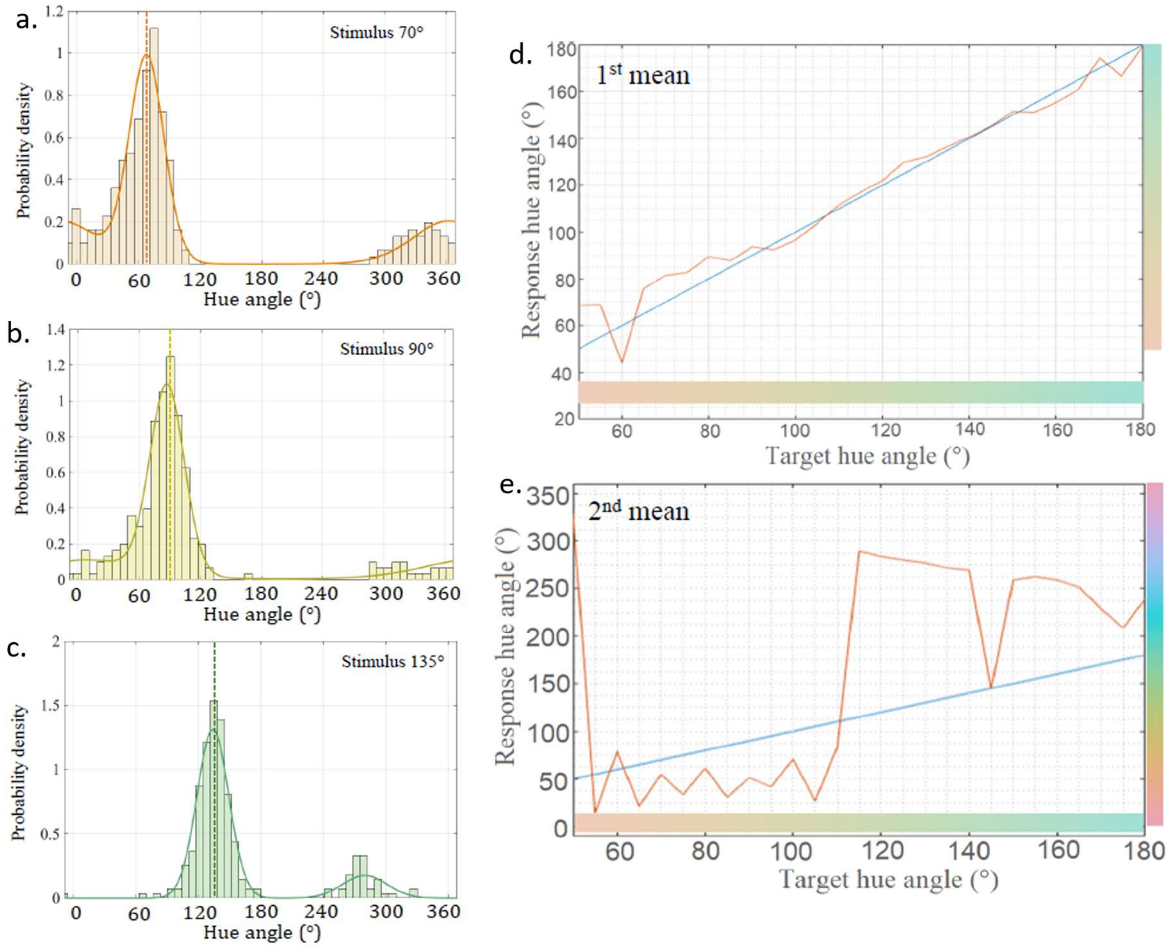


Figure 2.4 *Responses to the colour estimation task fit with a von Mises mixture model. (a-c) Histograms showing probability density across the hue circle for three target stimuli: a - 70°, b - 90°, c - 135°. The line shows the fit of the von Mises mixture model to stimulus specific responses, the vertical dashed line shows the hue angle of the stimulus, and the colour of the bars demonstrates the colour of the target stimulus. (d-e) The target hue plotted against average response hue for the first (d) and second (d) mean of each von Mises distribution (orange line). Target hue angle varies from 50°-180°, whereas response hue angles are distributed across the entire hue circle (360°). The blue line shows y=x, indicating where the target is equal to average response. The closer the orange line is to the blue line, the more accurate the responses are. The colour bars visualise the target and response hue angle.*

2.4 Discussion

Three main findings were observed: 1) response frequency was not uniformly distributed across stimuli; 2) responses were biased towards 2 category foci: yellow-orange and green, and away from the boundary between yellow and green; 3) there was a peak in response frequency for blue colours, despite not presenting any blue stimuli to participants. Therefore, the categorical biases observed by Bae et al. (2015) were replicated with different participants and lab equipment: response frequencies were greater at category foci and reduced at category boundaries. The peak in blue responses was an additional, interesting result.

The colour specific biases observed here are noisier than those by Bae et al. (2015). This is possibly due to training effects, as three participants completed 3600 trials each in Bae et al. (2015) study, compared with 15 participants completing 432 trials each in this replication. As such, the participants in the current experiment were less practiced and therefore could be more prone to making errors. In the current study, there was a very small but significant effect of trial number on response accuracy (r = -.075, n = 6480, $p <$ .001). Therefore, it is possible that responses were on average more accurate across 3600 trials compared with 432 trials.

Discrete category foci for orange and yellow were not observed in the present study. Bae et al. (2015) also did not observe a large peak in response frequency for the yellow category focus, although there is still a small distinction between the orange and yellow category. According to the category naming study in Bae et al. (2015) the yellow category is quite narrow, spanning across a 36° hue angle in CIELAB space compared with orange which covers 56°. In addition, the colour space was sampled more coarsely in this replication study, with 5° increments between stimuli compared to the original study which used 2° increments. As such, there were only seven stimuli in the present study which belong to the yellow category, which may not be enough to identify clear category boundaries and foci, thus explaining why a joint category for orange and yellow was observed here.

Another difference between the replication and the original findings is the systematic error in participants' responses to green stimuli. Curiously, participants incorrectly estimated green stimuli as blue, despite no blue stimuli being present in the sample. This finding was not identified by Bae et al. (2015), as they sampled stimuli from the entire colour wheel. As such this systematic error would not be apparent in their data unless response frequencies for individual stimuli were investigated. There was also a systematic error in responses to orange colours as pink (Figure 2.4a), however this error was not as large, as can be seen by the difference between the blue and orange lines in

Figure 2.4e. In addition, response frequency gradually declined from orange to pink, whereas there is a trough in responses at the boundary between green and blue, showing a clear category boundary that responses were biased away from.

To explain their findings, Bae et al. (2015) proposed the dual content model of colour working memory. This model combines a continuous estimate distributed around the noisy stimulus, with a categorical estimate distributed around the focus of the category assigned to the stimulus. The product of these two distributions is taken, giving an unconscious inference about the most probable colour. Category assignment is noisy, and thus influenced by uncertainty. For example, colours along a category boundary have more variable category assignment, so identical stimuli can be assigned to different colour categories on independent occasions.

In the current experiment, the systematic error of blue responses to green stimuli can be explained by the dual content model. Participants may be incorrectly assigning green stimuli to the blue category under high uncertainty. The blue categorical distribution therefore biases the noisy continuous distribution towards the blue category focus and away from the blue-green boundary. The results cannot be explained by a model which only accounts for noisy, continuous estimates of colour.

Errors in responses to orange stimuli as pink can also be explained by the dual content model, however the boundary between orange and pink responses was noisier than for blue and green. This may be explained by the warm-cool asymmetry in colour naming. Languages universally tend to have more colours to describe warm colours than cool colours (Gibson et al., 2017), and in the hue circle sampled, there are more colour terms in the warm region (yellow, orange and pink) than the cool region (green and blue). As a result, the category foci for warm colours are closer together than for cool colours, leading to noisier category boundaries. The fact that some category boundaries are noisier than others also demonstrates the importance of measuring responses across multiple category boundaries in order to fully understand the high level representation of colour.

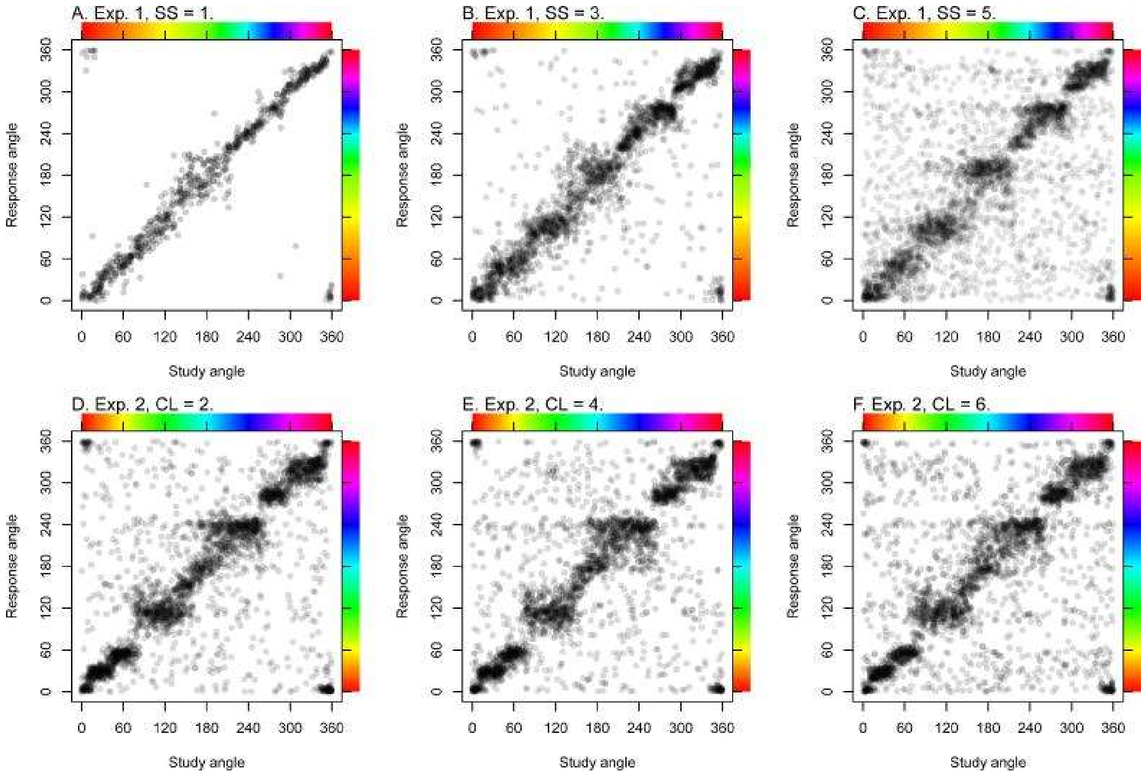


Figure 2.5 ***Scatterplots of data from Hardman et al. (2017) showing response hue angle plotted against study angle. A continuous response will follow the line y=x with some noise. Increasing set size (SS - top row) and cognitive load (CL - bottom row) creates a more discrete staircase effect, demonstrating categorical colour representation. Reprinted from "Categorical working memory representations are used in delayed estimation of continuous colors." By K. O. Hardman, E. Vergauwe, T. J. Ricker, 2017, Journal of Experimental Psychology: Human Perception and Performance, 43(1), pp.30-54. Copyright 2020 by APA.***

Hardman, Vergauwe and Ricker (2017) found further evidence to support the dual content model in visual working memory by identifying a staircase pattern when target and response hues are plotted against each other. The steps indicate clustering of responses around category foci. As set size or cognitive load was increased, the steps became more discrete, indicating an increased use of categorical distributions to estimate target colour under greater uncertainty (Figure 2.5). Hardman et al. (2017) extended the dual content model to allow flexibility in the proportion of continuous and categorical information used in visual working memory, depending on task demands. As set size or cognitive load increases, uncertainty increases, and participants rely more heavily on categorical estimates, leading to more discrete steps in response bias.

According to the model by Hardman et al. (2017), if participants' responses were biased by category structure, then a staircase pattern in average response bias would be predicted. In the present study, the plot of the 1st mean of the von Mises mixture model

(Figure 2.4d) against target hues is fairly linear, demonstrating that average responses are close to target values. In a similar study, Zhang and Luck (2008) produced a heatmap for response hues given a single target hue, and also observed a linear relationship, which was used as evidence against categorical biases in colour perception.

Bae et al. (2015) replicated the analysis by Zhang and Luck (2008) with their delayed colour estimation data, and also did not observe a staircase pattern in the relationship between target and response hues, despite observing categorical biases in response frequencies. Bae et al. (2015) argued that a linear relationship between target and response hues does not disprove the dual content model of visual working memory, as noisy category boundaries may not always be identified using the staircase method. Therefore, it is not surprising that a continuous distribution was observed in this analysis rather than discrete steps.

In addition, Hardman et al. (2017) only found a staircase pattern by increasing set size and memory load. The staircase pattern was not visible in their dataset for a set size of 1 and a 1500ms delay. This suggests the absence of a staircase pattern observed in many studies investigating categorical biases in colour working memory, may simply be due to cognitive load not being high enough to detect the categorical effect using an analysis of average response bias. However, as demonstrated in the current study, evidence for categorical biases can still be observed in tasks with less cognitive load by analysing response frequencies across hues. Furthermore, fitting the data with a mixture model and plotting the second mean response against target hues demonstrates categorical biases in average responses, despite low cognitive load (Figure 2.4e).In summary, the current replication study provides further support for categorical biases in colour representation. These biases may be even more pronounced with a larger cognitive load. When communicating about colours in the environment, one is generally required to hold colour representations in memory for longer than 900ms, so it would be relevant to investigate how a longer delay influences colour representation. In addition, colours in the environment vary much more widely than those used in previous experiments. It would be interesting to measure how categorical biases vary across the entire colour space. This is a challenge due to the sheer number of stimuli required to finely sample the space, and due to the difficulty of presenting multidimensional colour spaces to participants.

3 The geometry of perceptual colour space.

There is evidence for categorical biases in colour perception across the hue dimension (e.g. Bae et al., 2015; Persaud & Hemmer, 2014). However, by limiting colour spaces to one dimension, the results cannot be generalised to colour representation of all visible colours. Here, the properties of a high-level representation of colour are estimated by testing colour memory across hue and saturation in CIE1931 colour space for two delay lengths: 900ms and 3900ms. Categorical biases were observed across hue and saturation, corresponding to BCTs: green, orange, red, pink, blue and grey. Responses were also biased towards more saturated colours, with a stronger effect in the long delay condition. It is proposed that these biases are due to a non-uniform prior over colours which directs the visual system towards perceiving and remembering the most informative colours under conditions of uncertainty.

3.1 Introduction

In Chapter 2, the colour estimation experiment in Bae et al. (2015) was reviewed and replicated. In their experiment, Bae et al. (2015) sampled stimuli varying in hue in CIELAB space, spanning across six BCTs: blue, pink, green, purple, orange and yellow. They observed that participants' responses were biased towards the focal points of these six categories and away from category boundaries.

Since Bae et al. (2015) demonstrated the importance of measuring stimulus specific effects in colour perception and memory, many other researchers have replicated these findings and thus found further evidence for a categorical prior in colour perception (Cibelli et al., 2016; Donkin et al., 2015; Hardman et al., 2017; Hu et al., 2014; Persaud & Hemmer, 2014). In addition, some studies have varied cognitive load to measure the influence of uncertainty on categorical biases in visual working memory. For example, Hardman et al. (2017) conducted two delayed colour estimation experiments. In the first experiment, set size was varied between 1, 3 or 5 target colours with a 1500ms delay. In the second experiment, memory load was manipulated using a secondary auditory discrimination task during a 6s maintenance interval. As set size and memory load were increased, working memory accuracy decreased, and responses became more categorical. These findings support a model of visual working memory where categorical information may facilitate noisy continuous representations.

Donkin et al. (2015) aimed to explicitly measure the role of colour categorisation in colour memory. They conducted standard delayed colour estimation trials and alternative labelling trials. In the alternative labelling trials, participants were asked to free recall the colour category of a target. Three trials later, they were asked to reproduce the continuous value of the target stimulus using the self-reported label as a cue (Figure 3.1). The assumption was that after three trials, the perceptual memory for the target colour would have decayed, so responses would be determined by labelling processes alone. Therefore, responses to labelling trials produced a purely categorical estimate of the target colour. Responses to standard trials were then modelled as a mixture of three distributions: a continuous estimate, a categorical estimate using the verbal label, and noise from a random guess. Including a categorical estimate of the target colour significantly improved the fit of the model compared with just a continuous estimate, further suggesting that categorical information is used in visual working memory.

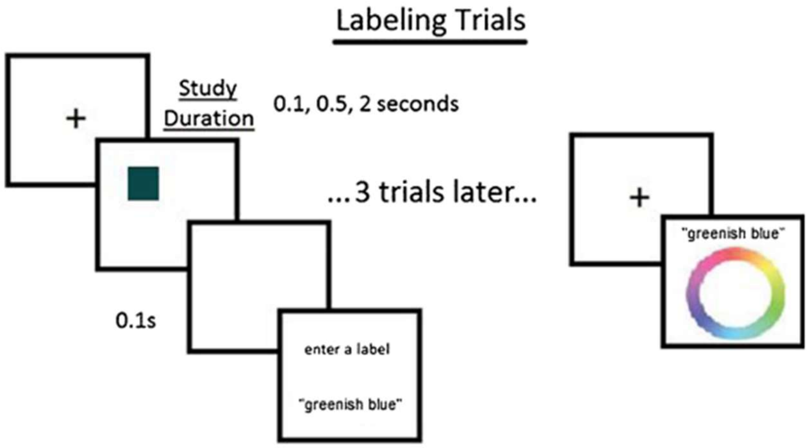


Figure 3.1 *An illustration of the labelling trials in Donkin et al. (2015). Reprinted from "Verbal labeling, gradual decay, and sudden death in visual short-term memory." By C. Donkin, R. Nosofsky, J. Gold, and R. Shiffrin, 2015, Psychonomic Bulletin & Review, 22(1), pp. 170-178. Copyright 2014 by the Psychonomic Society, Inc.*

While plenty of research has been conducted on colours varying in hue, the saturation dimension has been broadly neglected, and in some cases not even kept constant. The advantage of using colours varying in only one dimension is that the hue circle provides an intuitive way of presenting the continuous colour scale to participants, and responses are much simpler to analyse, as response bias can be measured using the angular difference in the hue circle. However, since colours fundamentally vary in multiple dimensions, the results cannot be used to make inferences about the geometry of high-level colour representation.

By using a limited colour space to collect responses, it could be argued that biases occur in the response rather than in the representation of colour. For example, if a participant was presented with a pale pink colour, their representation could drift towards a more saturated hot pink colour during the delay. Hot pink is not an option in the response hue circle, and therefore the participant would have to pick the best possible option from the limited choice. By confining responses to a limited sample of colours at a single saturation level, it is possible that participants are forced to create boundaries and draw distinctions that are not an element of their perceptual colour space, but rather a feature of the colour space used in the experiment.

Additionally, in a two-dimensional colour space, as the distance from central grey increases towards more saturated colours, the distance between category foci increases, and it is easier to discriminate across category boundaries. Bae et al. (2015) and Donkin et al. (2015) used a hue circle in CIELAB space with a radius of 38 and 40 respectively. At these medium saturation levels, the distance between category foci is smaller, and therefore responses may be more prone to errors compared with responses given on an RGB triangle varying in hue and saturation.

Saturation is an important dimension of colour, because it discriminates between achromatic and chromatic colour categories. Jameson and Alverado (2003) found evidence that terms discriminating lightness and saturation of colours are more consistent across languages than terms discriminating between hues. Comparing English and Vietnamese language groups, there were differences in the mappings of BCTs which mostly vary in hue, but modifier terms used to modify the lightness and saturation of BCTs (such as light, dark, vivid, pale) had similar mappings between the languages.

As such, Jameson (2005) emphasised the importance of lightness and saturation dimensions when proposing their interpoint distance model of colour term evolution. In this model, the irregularly shaped perceptual colour space is sequentially partitioned into categories which maximise the difference between them and maximise similarity within them. This universal constraint on colour naming minimises individual and interpersonal uncertainty about colours during perception, memory and communication, and explains

Berlin & Kay’s sequence of evolution of BCTs (Berlin & Kay, 1991). Jameson (2005) proposes that the black/white classification in the first stage polarises the lightness dimension, and the red/yellow classification in the second stage polarises the saturation dimension. As such, they argue that brightness and saturation are more important than hue for colour categorisation, as they impose the earliest partitioning of the colour space.

Regier et al. (2007) tested the interpoint distance model using data from the World Colour Survey (WCS; Cook et al., 2005). They organised Munsell colour chips in CIELAB colour space to measure the perceptual distance between colours. The colour chips were then artificially assigned to categories to generate an optimal partitioning of the space, such that distance between colours within a category is minimised, and distance between colours across categories is maximised. Regier et al. (2007) found that most languages in the WCS had similar colour naming maps to the optimal partitioning of the space, such that category foci tended to be maximally distant from each other in the CIELAB space. Consequently, focal colours were situated around the edge of the colour space. As the centre of CIELAB space is achromatic and the edges have high saturation, this means that focal colours were more saturated than intermediate colours.

Witzel et al. (2015) also identified a significant positive correlation between focal colours from the WCS and Munsell chroma, where Munsell chips with the highest saturation were more likely to be selected as category prototypes by participants (Figure 3.2). Witzel (2018) pointed out that influential studies on colour naming used maximally saturated Munsell chips (e.g. Berlin & Kay, 1991; Lindsey et al., 2015; Regier et al., 2005), which varies strongly across the Munsell space. As such, variation in Munsell chroma across stimuli may have biased participant’s responses in colour naming studies, thus highlighting the importance of controlling saturation in the stimuli.

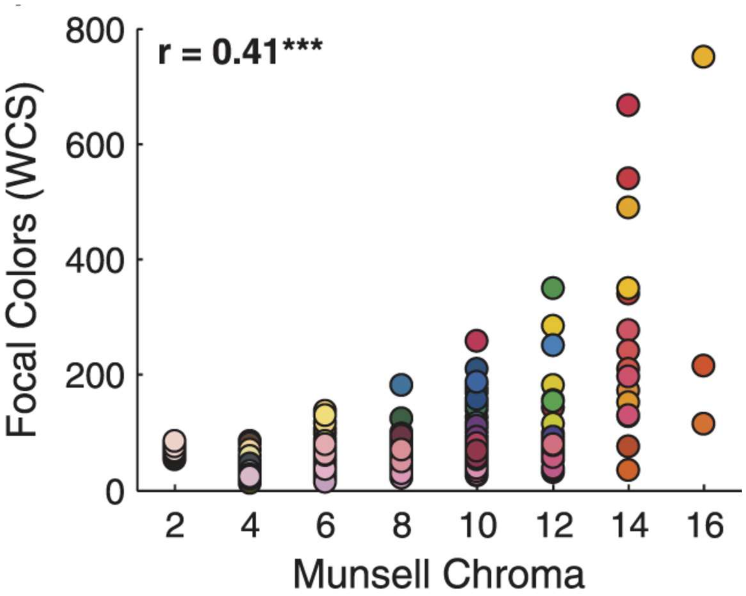


Figure 3.2 *Correlation between focal colours and Munsell chroma across Munsell chips (Witzel et al., 2015).Focal colours refer to the number of prototype choices in the WCS (Regier et al., 2005). Each coloured circle represents a Munsell colour chip. Correlation coefficient is given in the top left hand corner and is significant (p <.001). Reprinted from "What determines the relationship between color naming, unique hues, and sensory singularities: Illuminations, surfaces, or photoreceptors?" By C. Witzel, F. Cinotti, and J. K. O'Regan, 2015, Journal of Vision, 15(8):19, pp.1-32. Copyright 2015 ARVO.*

Witzel (2016) reanalysed the data from Lindsey et al. (2015) to test whether saturation variability in the stimuli confounded their results. Lindsey et al. (2015) investigated the evolution of colour terms in different languages. A high consensus colour term is given the same categorical label by the majority of speakers in a language. They found that a non-industrialised small scale society had only three high consensus colour terms. However, low consensus colour terms across the community were grouped similarly to the BCTs, showing a collective use of colour terms with a similar structure to universal colour naming patterns. However, Witzel (2016) pointed out that saturation was not controlled in their stimuli. When reanalysing the data, he found that colours which were high or low in saturation had higher naming consensus than those with median saturation, thus explaining the grouping of low consensus terms in this study.

Clearly saturation plays an important role in colour categorisation, but it has been neglected in studies investigating categorical biases in colour perception. As such, extending colour estimation tasks into the saturation dimension is an important step towards achieving an accurate mapping of the geometry of perceptual colour space. Since participants are able to browse and respond with any visible colour, the responses can be used to accurately estimate the non-uniform prior which biases colour representation towards the most likely outcomes.

This is the motivation for the current experiment, where the hue circle from the first experiment is replaced with an RGB triangle in CIE1931 xyY chromaticity space. This space was chosen because it spans across six basic colour terms (green, orange, red, pink, blue and grey) varying in hue and saturation, and it is a widely used standard that allows comparison to other spaces. While using a triangle rather than a circle may introduce some biases near the corners (J. D. Mollon, 2009), it was deemed beneficial to use a triangle in order to include foci of red, green and blue categories. In addition, the triangle is roughly equilateral, so the bias is assumed to be similar across the three corners. Stimuli are presented on a ring around the RGB triangle, enabling both stimuli and responses to vary across hue and saturation dimensions.

In addition to using a different colour space, delay length was also manipulated. Previous studies controlled cognitive load by varying the set size, the duration stimuli were presented for, or the delay between target and response. Increasing cognitive load leads to more uncertainty in continuous representations of the target colour, which results in reduced response accuracy and a heavier reliance on categorical estimates (Bae et al., 2015; Donkin et al., 2015; Hardman et al., 2017). Therefore, it was predicted that increasing delay length would result in more uncertainty and larger categorical biases in responses.

The aim was to identify biases in colour perception across hue and saturation dimensions for two delay lengths, 900ms and 3900ms. Response bias was measured as the variability between target colour and average response. Responses were predicted to be biased towards colour categories and away from category boundaries, corresponding to BCTs. The response error (distance between target and average response) was also predicted to be larger for the longer delay condition due to increased uncertainty.

3.2 Methods

3.2.1 Participants

Thirty-one participants were recruited, of which 22 were female. Thirty participants were University of Bristol undergraduates who received course credit for participation.

One participant was a volunteer recruited from outside the University. All participants had normal or corrected to normal vision, checked using the Ishihara test for colour vision deficiency. Protocol was approved by the School of Psychological Science Research Ethics Committee.

3.2.2 Equipment

The experiment was conducted in a dark, quiet room where the monitor was the only light source. Stimuli were presented on a ViewPixx/3D CRT monitor, with a spatial resolution of 1920 x 1080 pixels, which had been calibrated according to standard methods (Hansen & Gegenfurtner, 2013) using a Minolta CS-100 photometer. Participants sat at a viewing distance of 60cm with a fixed chin rest. A standard computer mouse was used as an input device. The display subtended 47.42° by 27.75° of visual angle.

3.2.3 Stimuli

Eighty stimuli were selected, equally spaced across the CIE1931 chromaticity space. Stimuli varied in hue and saturation, and luminance factor for Y was 50%. Stimuli were presented on a large ring over a uniform grey background. The perimeter of the ring was 12.39° from the centre of the display, and 3.14° wide.

3.2.4 Procedure

Each trial began with a white fixation cross (1.57° x 1.57°) displayed in the centre of the screen for 500ms. Following fixation, the stimulus was presented for 100ms. After a delay, an RGB triangle (13.1° x 13.6° x 14.9°) was displayed, which participants were able to browse. As they browsed the triangle using the computer mouse, the colour of the ring varied, and participants clicked the mouse when they were satisfied with their response colour. Participants were not restricted in how long they had to respond. The RGB triangle rotated at random between blocks to control for motor biases from associations between colour and position. The mouse was set to return to the top of the screen at the start of each trial, ensuring that the coloured ring was grey before participants moved the mouse.

There were two delay lengths, 900ms and 3900ms. Each participant completed four blocks of 40 trials for each delay condition, therefore observing each stimulus twice per condition. As such, across all participants, 62 observations were collected per stimulus per condition, which is greater than the number of measurements obtained in previous delayed estimation experiments (Bae et al., 2015). Stimuli were presented at random, and participants completed each condition one after another, in a counterbalanced order. There was a break every 40 trials.

After 10 practice trials, participants were informed that their performance was being recorded and they were required to perform above a certain threshold. This ensured that participants remained attentive throughout the experiment.

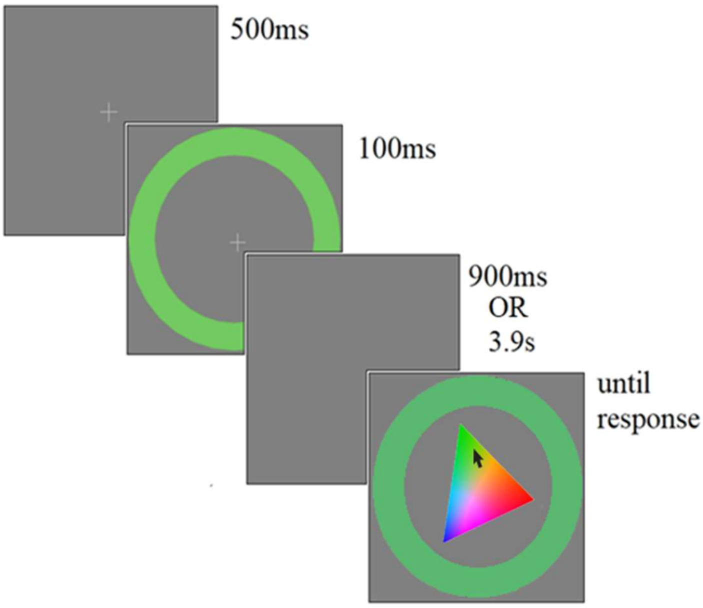


Figure 3.3 *Procedure for delayed colour estimation task across hue and saturation dimensions.*

3.3 Results

All responses were transformed from RGB to CIE1931 xyY chromaticity coordinates using colour transformation functions (Lindbloom, 2017). The Y parameter was dropped, and xy chromaticity values were averaged for each stimulus, across all participants. Any responses above three standard deviations from the mean stimulus-specific response were removed.

Table 3.1 shows descriptive statistics for reaction times in each delay condition. On average, reaction times are longer for the long delay condition. Figure 3.4 a-b shows average reaction times across the stimuli for each condition. The desaturated, grey stimulus at (0.33,0.33) has a notably longer reaction time compared with all other stimuli.

Figure 4c-d shows the distribution of responses for each target hue. An ellipse was fitted to each distribution using the eigenvectors of the covariance matrix as the major and minor axes. Here, they have been scaled down to 50% of the confidence interval so that the ellipses do not overlap too much. Response distributions are not equally dispersed across different colours: some colours have larger variation.

Table 3.1 *Descriptive statistics for reaction times in seconds.*

	Min	**Max**	**Mean**	**Median**
Short delay condition	0.51	34.7	3.64	2.95
Long delay condition	0.56	62.9	5.16	4.06

Systematic biases in responses across the colour space were displayed in a vector graph, where an arrow is plotted from the target colour to the average response (Figure 3.5). These arrows illustrate the size and direction of average response bias for each stimulus. Stimulus-specific biases are observed, where the size and direction of bias varies across colours. Variation in responses is unrelated to the MacAdam ellipses (Appendix A).

Vectors around the edge of the triangle tend to be larger and biased towards the centre of the triangle. These stimulus specific biases are due to edge effects, where responses to stimuli around the edge of the triangle are restricted in the direction that they can be biased in. This results in a larger effect size in a specific direction for stimuli on the edge.

Patterns in participants' responses are also observed on the inside of the triangle, where responses appear to be biased in a categorical structure. There are six focal points, roughly corresponding to six BCTs; green, orange, red, pink, blue and grey (Figure 3.5c). The category foci function as attractors in the colour space as responses to colours nearby are biased towards them. These patterns in responses were observed in both the short and long delay conditions, as demonstrated by very similar patterns of bias (Figure 3.5c). Therefore, the categorical biases observed are a stable effect in colour perception across hue and saturation.

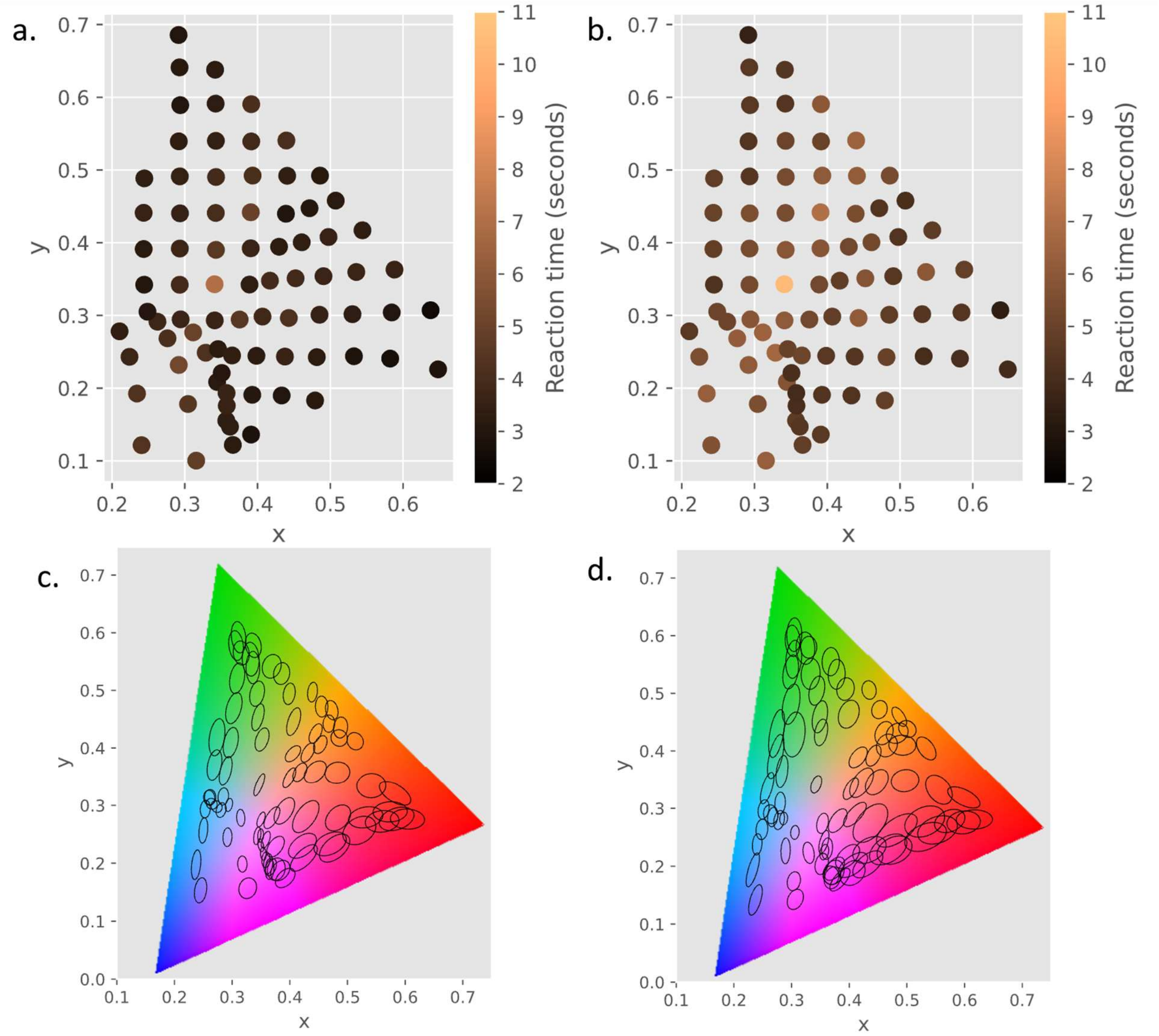


Figure 3.4. *Reaction times and covariances across target stimuli. (a-b) Average reaction time across stimuli for short delay condition (a) and long delay condition (b). (c-d) Distribution of responses for specific stimuli in the short delay condition (c) and long delay condition (d). Ellipses are fit to covariance matrices at 50% confidence interval.*

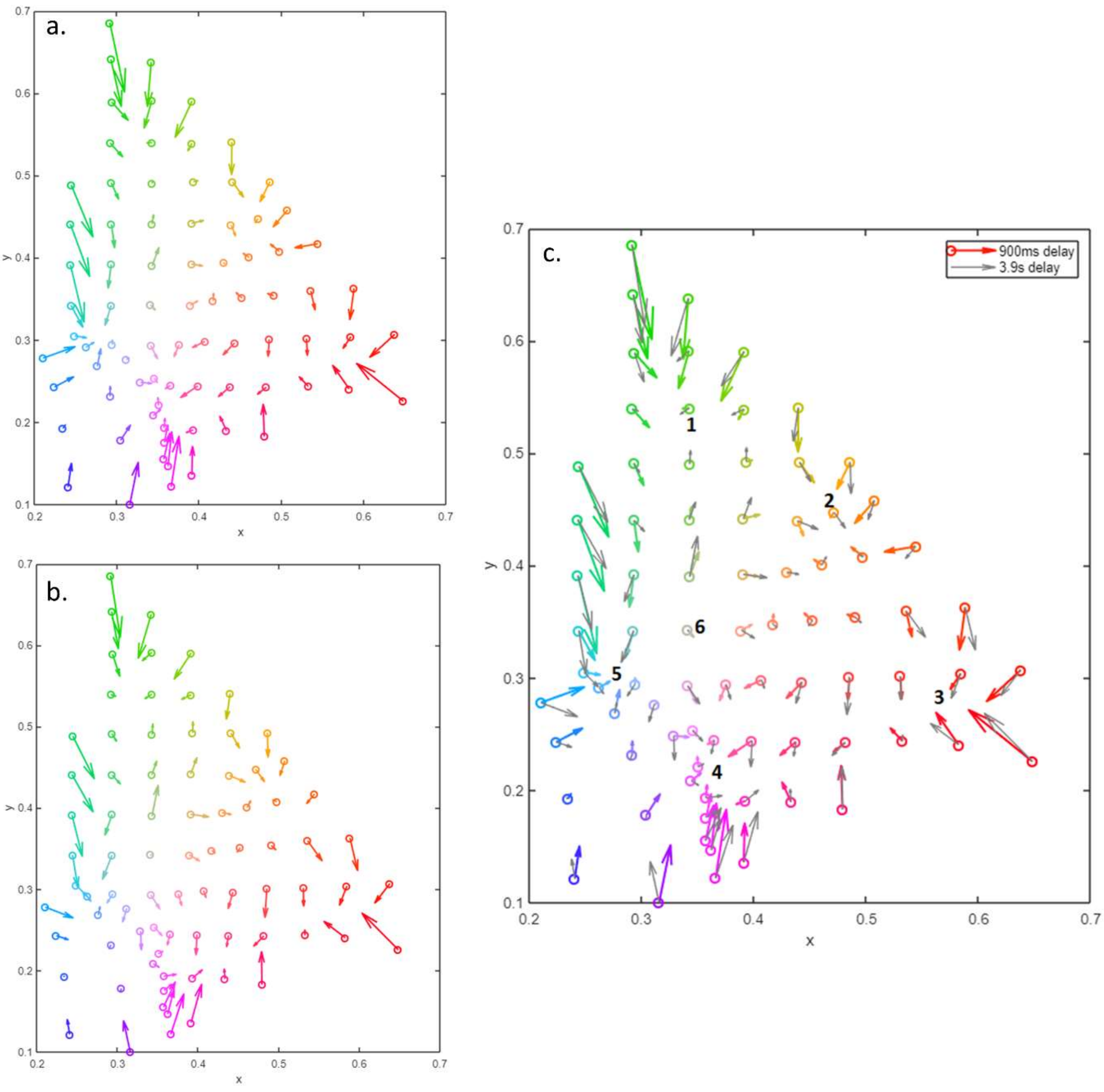


Figure 3.5 *Vector field showing average response bias in CIE1931 chromaticity space for 900ms delay (a), 3.9s delay (b) and both delay conditions (c). The numbers in (c) correspond to the six category foci in the space: 1-Green, 2-Orange, 3-Red, 4-Pink, 5-Blue, 6-Grey.*

It was predicted that uncertainty would be greater for the longer delay condition than the short delay. Response error was measured as the distance between target colour and average response, where a larger response error would suggest more perceptual uncertainty. Response error was measured for each stimulus and condition. Figure 3.6 shows a line of best fit for response errors in the short delay condition, plotted against response errors in the long delay condition. The line of best fit is similar to *y*=*x*, suggesting that response error does not differ between conditions. A paired-samples t-test confirmed that there was no significant difference in size of response error between conditions (t = 1.60, p = 0.114). This suggests that uncertainty was not significantly greater when delay length was increased from 900ms to 3900ms.

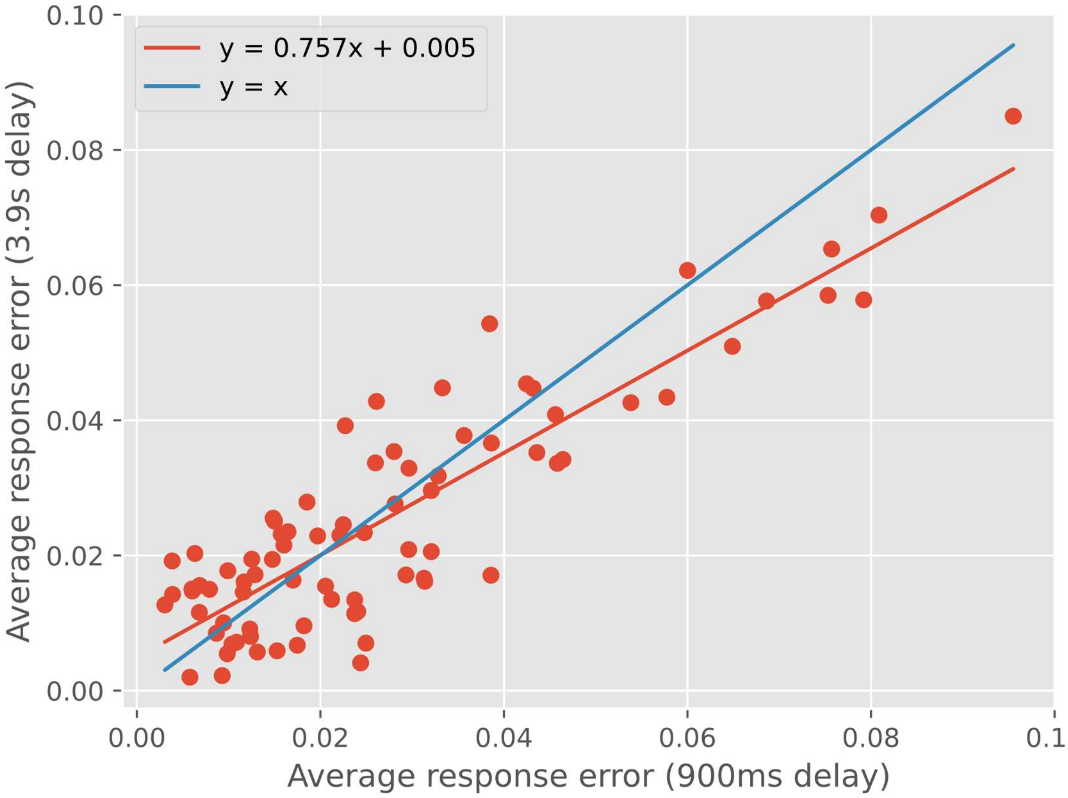


Figure 3.6 *Scatterplot showing average response error for 80 stimuli in short delay condition (900ms) vs long delay condition (3.9s) for a colour estimation task. The red line shows the line of best fit using least squares regression. The blue line shows y=x.*

By extending the colour estimation task to measure colours varying in saturation as well as hue, the size of the colour space was greatly expanded. For a behavioural experiment, this makes it difficult to sample the space very finely, as increasing the number of stimuli requires spending more time testing participants, which is resource intensive. However, coarsely sampling the space can be problematic, as it makes it more difficult to identify category boundaries.

The coarse sampling issue was resolved using kernel smoothing to estimate data as a weighted average of neighbouring observed data. This produced an estimation of participants' responses to colours in smaller steps across the colour space. A Gaussian kernel smoother was used to determine the weight applied to neighbouring data. Kernel sizes were tested between 0.01 and 0.08 to determine the optimal trade-off between bias and variance (Appendix B). Smaller kernels have high variance and do not vary smoothly, whereas larger kernels are very smooth but have high bias so some of the signal is lost. A kernel size of 0.04 was selected to produce a smooth output which captured the true signal.

Once estimates of the data sampled evenly and finely across the colour space were produced, the response biases were again plotted on a vector field from target to average response estimate (Figure 3.9a-b). The attractors are more apparent for the smoothed data.

To test how responses varied as delay length increased, the response bias in the short delay condition was subtracted from response bias in the long delay condition (Figure 3.7). Responses move away from the achromatic centre and towards saturated colours as the delay length increases. To test whether this bias is significant, the difference between the unsaturated grey point at (0.33,0.33) and the average response was calculated for each stimulus in each condition.

A sign test was then conducted to measure if there was a significant difference in the median distance from grey between average responses in the long delay condition compared with the short delay condition. The distance of the average response from grey was significantly larger in the long delay condition (Z = -4.89, p < .001), indicating a significant bias toward more saturated responses as delay length is increased.

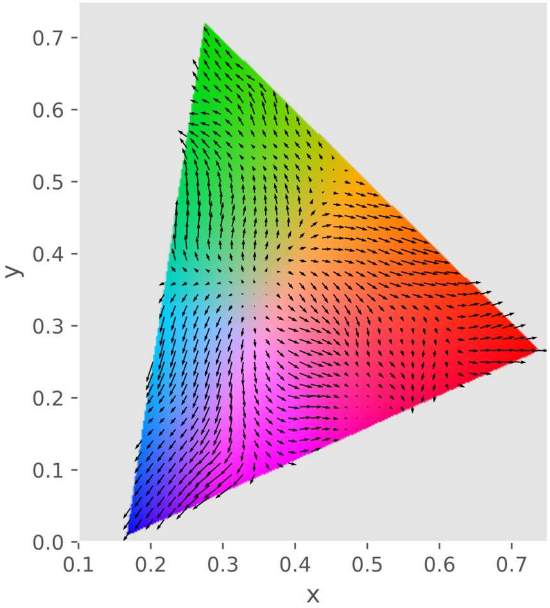


Figure 3.7 *Difference in kernel smoothed responses as delay length increases from 900ms to 3900ms.*

To formalise the precise regions of category foci and boundaries in the kernel smoothed responses, the vector field was treated as a fluid flow where vectors represent the direction and rate of flow. Divergence was calculated across the space, which transforms a vector into a scalar value by measuring the change in density at each point. A source is an area from which fluid flows, and has a divergence greater than zero as density is decreasing (Figure 3.8). A sink is an area towards which fluid flows, and has a divergence less than zero as density is increasing. Areas with an equal amount of flow in and out will have a divergence of zero, as density is not changing.

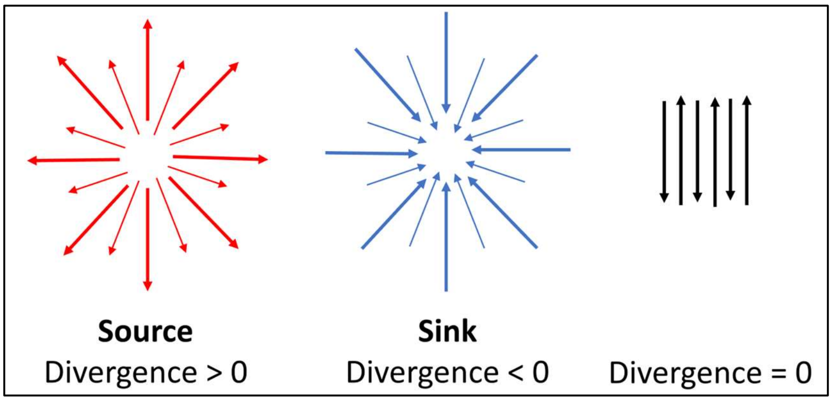


Figure 3.8 *Fluid flow and divergence ranges of sources and sinks. (see text for details)*

A contour plot shows the divergence of response biases across the chromaticity space (Figure 3.9c-d). The sinks in the colour space show regions towards which responses are biased, thus indicating the presence of category foci. The saturated colour categories are sinks and the achromatic grey category is a source, with and additional source on the boundary between red and orange. The sinks and sources are very similar across delay conditions, although the divergence of the grey source is larger in the long delay condition, reflecting the bias towards more saturated colours. There are sinks for the following BCTs: green, orange, red, pink and blue.

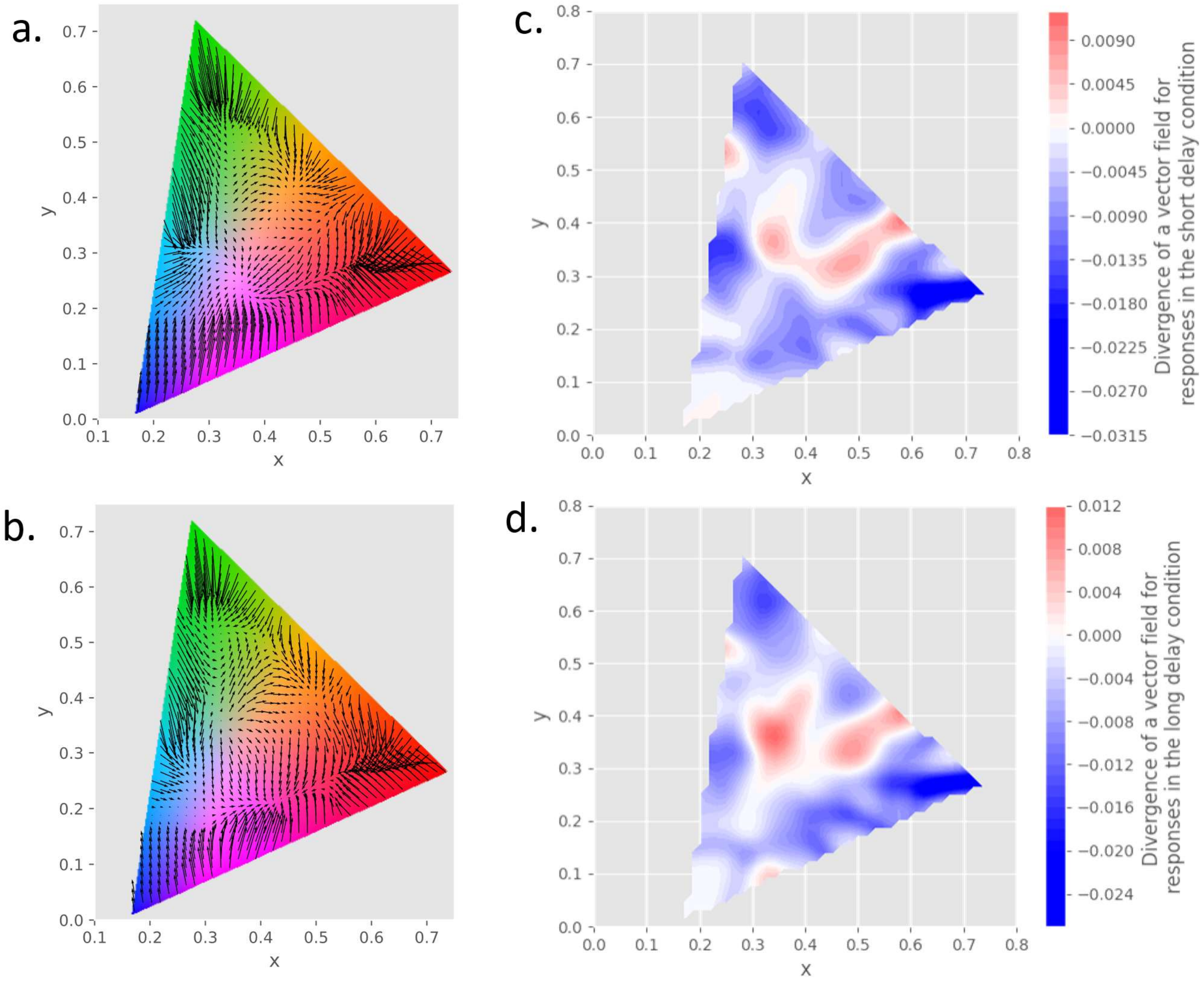


Figure 3.9 *Vector fields showing kernel smoothed average response bias and divergence of the vector field. (a-b) Kernel smoothed average response bias in a colour estimation task for 900ms delay (a) and 3.9s delay (b). (c-d) Contour plot showing divergence of smoothed average response bias for 900ms delay (c) and 3.9s delay (d). The blue regions are sinks, and the red regions are sources.*

3.4 Discussion

In a delayed colour estimation task, categorical biases in colour representation were observed across hue and saturation dimensions for two delay lengths. Response variability was not uniform across the space, but biased towards six focal points corresponding to BCTs: green, orange, red, pink, blue and grey. These categorical effects were stable across both delay conditions. The category foci vary with saturation, where unsaturated colour categories are sources that repel responses, and saturated colour categories are sinks that attract responses. There was no significant difference in response error between the two conditions, although response bias was significantly more saturated in the long delay condition.

These biases were observed by calculating average response error. Hardman et al. (2017) warned against the modelling of response error in colour estimation tasks for one-dimensional colour spaces, as category boundaries are probabilistic, and therefore biases in responses may be averaged out. For example, at a category boundary in a hue circle, if 50% of responses are biased clockwise, and 50% are biased anticlockwise, then the average response error will be 0, suggesting that responses at the boundary are highly accurate, which is not the case. To avoid this error, Bae et al. (2015) measured response frequency, and Hardman et al. (2017) plotted target and response angles against each other on a scatterplot. However, with a two-dimensional colour space, it is not possible to plot target and response colours on a two-dimensional scatter plot. Additionally, due to non-linear colour transformations, the distribution of target colours was not uniform across the colour space, so response frequency is also not a valid measure in this study. Finely sampling the two-dimensional colour space using kernel smoothing reduces the risk of misinterpreting a category boundary as a category focus, as surrounding vectors will demonstrate whether responses move towards or away from a region, further measured using divergence.

3.4.1 Effect of delay length on uncertainty

Previous studies reported that increasing memory demands, for example delay length, increased the uncertainty of responses, and therefore produced a larger categorical effect (Bae et al., 2015; Donkin et al., 2015; Hardman et al., 2017). Such findings lend support to a Bayesian model of colour perception, whereby factors which increasethe uncertainty of sensory signals result in an increased reliance on prior knowledge to make conclusions about the signals. Therefore, under a Bayesian framework, a relationship between bias and uncertainty is predicted, and such an effect has been observed in other visual domains (Ashourian & Loewenstein, 2011; Girshick et al., 2011). However, in the present study, no significant difference in response error was observed between delay conditions, suggesting that there was no effect of delay length on uncertainty, despite observing a significant effect of delay length on saturation bias.

Even though these results seem surprising since memory generally increases the uncertainty of representations, other studies have also failed to find significant differences in responses for the delay lengths used here. Donkin et al. (2015) compared the distribution of responses for a colour estimation task with delay length varying between 0.1s and 10s. Response variation did increase as a function of delay length, however the difference between the variance for 1s and 4s delay lengths is minor. Similarly, Nemes et al. (2010) reported an effect of delay length from 0s to 10s on response variation for a colour discrimination task, but the difference between 1s and 5s delay was very slight. Olkkonnen et al. (2014) also observed that increasing delay length (internal noise) and chromatic variability of stimuli (external noise) had a significant effect on response bias, but no significant effect on the variability of responses.

As such, it is possible that the measurements were not sensitive enough to measure a difference in uncertainty for the two delay lengths tested here. Since this study used a large number of participants for a psychophysics experiment, if response variability differs between observers, then it may be difficult to detect a difference in variability between the two conditions when averaging variance across observers. Olkonnen et al. (2014) observed a small but significant positive correlation between response bias and variability for each individual observer, although no correlation was observed when results were averaged across observers. Therefore, the average response across 31 participants may not have been sensitive enough to measure differences in uncertainty between delay conditions.

Another potential explanation for not observing an increase in uncertainty is due to the edge effects. Using a triangle as opposed to a hue circle introduces edge effects, where responses are biased away from the edge of the triangle and towards its centre. This occurs because responses on the edge of the triangle are restricted in the direction they can be biased in. The effect reduces gradually as distance from the edge increases. In the long delay condition, responses are biased towards more saturated colours than in the short delay condition. This is the opposite direction of effect to the edge effects. However, these two effects are independent, as the saturation bias does not vary with distance from the edge of the triangle. If both of these effects are occurring simultaneously, then this could explain why response accuracy is not significantly different between the delay conditions, as the increased saturation bias in the long delay condition reduces the edge effects, thus making the average response closer to the target.

3.4.2 A non-uniform prior over perceptual colour space

Stimulus specific biases were detected in responses, suggesting that representations of colour space are not uniform. But the raw response data were noisier than in Bae et al. (2015). This is due to using a large number of participants and using a two-dimensional colour space. Bae et al. (2015) used a small number of highly trained participants which produces a clearer signal-noise ratio, but the results are not necessarily representative of how colours are maintained in working memory for everyday tasks. For example, Witzel and Gegenfurtner (2015) reported that trained observers did not demonstrate categorical perception in a speeded discrimination task, whereas naïve, inexperienced displayed better discrimination performance for colours across category boundaries. This suggests that trained observers do not perform similarly in perceptual experiments to untrained observers, and therefore the results of experiments using the former may not generalise to colour perception in everyday tasks.

In addition, previous studies used colours that varied in only one dimension, so the small colour space can be sampled finely to estimate perceptual representations using distance between target and response. By extending the colours into two dimensions, the colour space is much larger, and was therefore sampled more coarsely to ensure that enough observations are collected per stimulus in an experiment of reasonable length. In addition, responses vary in both distance and angular direction, and thus results are inevitably noisier to interpret.

To maximise signal-noise ratio, and investigate robust categorical effects across observers in a two-dimensional colour space, the data were processed using a kernel smoothing function. Divergence was then calculated to measure regions of the colour space which attract responses (sinks) and repel responses (sources). This is a novel method to be applied in colour perception, which produces a map of the geometry of high-level representations of colour. The contour plots of divergence are a very powerful and informative way of visualising the perceptual colour space and provide a route to a cohesive understanding of continuous colour representation. These novel methods can be applied to compare colour perception across language groups, to measure colour perception in people with colour vision deficiencies, or to investigate colour perception during stages of language acquisition to measure how learning additional colour terms influences perception. As such, it has a wide variety of applications with the potential to advance many areas of perceptual research.

The maps in Figure 3.9 represent the non-uniform prior which influences perception of colour space. Priors facilitate perception of noisy visual input by directing the observer away from generally unlikely conclusions and towards generally likely conclusions. These priors have stronger effects when the signal is noisier, for example when there is higher cognitive load or difficult viewing conditions. If the prior expectation of colours was uniform, then the divergence would be zero; however there are many peaks and troughs which roughly correspond with BCTs.

The sinks are the attractors: these are the conclusions which participants are drawn towards. The sources are regions where responses are repelled, so these conclusions are less likely. The current findings support a prior expectation to perceive more saturated and focal colours in working memory. In addition, increased memory demands in the long delay condition resulted in a significant shift towards more saturated responses compared with responses in the short delay condition. A longer delay between presentation and response increases cognitive demand, so participants relied more heavily on their prior to complete the task. This provides further evidence to support a prior which directs observers towards perceiving more saturated colours. Given that the experiment was very simple - participants were presented with a colour, which was estimated after a delay - it is striking that such robust effects were observed.

Previous studies have also reported a saturation bias in memory for colour matching tasks (Bloj et al., 2016; de Fez et al., 2001). Estimates of Munsell colour chips were more saturated than the target for simultaneous matching and a 10 minute delay, with a stronger effect for the delayed task (de Fez et al., 2001). This suggests that the prior towards saturated colours originates in perception, and has a greater effect under conditions of uncertainty.

However, Allred and Olkkonen (2015) found no evidence for a categorical or saturation bias in long term memory. Participants made colour matches to real objects under different illumination conditions and delay lengths, and adding a 10 minute delay to the colour matching task increased variability of responses (Allred & Olkkonen, 2015). There was no evidence for a response bias towards more saturated or more prototypical colours. Although these results do not align with the current findings, there are some differences between the experimental conditions which could explain the discrepancy. Allred and Olkkonen (2015) used two illumination conditions; 4000 Kelvins (K) and 2500K. A 4000K illuminant is natural daylight, however a 2500K illuminant is unnatural since it's warm coloured and highly saturated. These illumination conditions were relevant for testing colour constancy, but using a highly saturated illuminant is likely to bias responses. Considering that responses to both illumination conditions were averaged, this measure may not be sensitive enough to observe saturation biases in colour perception.

Furthermore, in a similar experiment, participants estimated the colour of objects under natural daylight, and representations were biased towards more saturated colours in short-term and long-term memory (Bloj et al., 2016). This suggests that a saturation bias does exist, but natural illumination is important for measuring a saturation bias in colour estimation.

3.4.3 The relationship between categorical bias and saturation bias

Colour categories vary with saturation, where category foci tend to be highly saturated for chromatic categories or unsaturated for achromatic categories. However, research examining whether focal colours are more saturated is mixed. Witzel (2016) identified an effect of saturation contrast on colour naming, where high and low saturation colours were easier to name than medium saturation colours. Colours which are easier to name are more likely to be category prototypes, and therefore closer to the category foci.

However, Witzel and Franklin (2014) found that prototypical red, yellow, green and blue are not more saturated than other colours in terms of discriminability or subjective appearance, therefore they concluded that focal colours do not appear more colourful. It is difficult to disentangle the relationship between colour categories and saturation, since it depends on the colour space. For example, Regier et al. (2007) found that focal colour chips in Munsell space stick out when plotted in CIELAB space. Colours which stick out are further from the central grey point, and therefore more saturated. As such, it is unclear whether high saturation of focal colours is a peculiarity of the colour space used or a feature of human colour perception.

3.4.4 What benefit is provided by a prior expectation to perceive category foci and saturated colours?

In general, a prior directs perception towards the most useful conclusion, to improve task performance under uncertainty. As such, the prior may direct observers towards saturated colours because these are the most behaviourally relevant colours. For instance, low saturation colours are more common than high saturation colours in the environment (Rosenthal et al., 2018; Yendrikhovskij, 2001), therefore high saturation colours may provide a more informative signal.

According to rate distortion theory, some errors in colour perception are more costly than others, and so a prior guides cognition towards making judgements which reduce the cost of misremembering (Sims et al., 2016). This would suggest that the cost of misremembering a saturated or focal colour is high because these colours are more behaviourally relevant; therefore humans are biased towards perceiving saturated, focal colours to reduce the likelihood of missing a potentially important signal.

However, a criticism of Bayesian models of cognition is that they use circular reasoning to explain processing, whereby a predictive model is generated from observed behaviour, and this is used to explain the behaviour. In the case of colour perception, categorical biases in colour perception are measured, and the categories identified in perception are used as evidence for categorical priors. However that doesn’t explain why the categorical priors exist and how they actually reduce uncertainty. Sims et al. (2016) and Bae et al. (2015) used different Bayesian models to successfully predict the categorical biases observed in the same dataset, however neither of these approaches tested whether the categorical prior provided an advantage to perceptual processes in a real task. It is necessary to test whether colour categories are useful in behaviourally relevant tasks in order to demonstrate that they contribute towards optimal cognitive processing during perceptual uncertainty.

Zaslavsky et al. (2019) investigated whether categorical priors may exist in the distribution of colours in the environment. Under the assumption that objects are communicated more often than backgrounds, Zaslavsky et al. (2019) modelled a prior distribution over colours as the probability of a colour appearing in a foreground object as opposed to the background of a natural image. The prior was significantly correlated with the distance of each colour from the central grey in CIELAB space, whereby the colours of objects in the foreground were further away from central grey than colours in the background. In CIELAB space, the central grey is the most unsaturated point in the colour space and, as distance from this point increases, colours become more saturated. As such, the colours of foreground objects were more saturated than background colours in natural images. This may explain why the prior directs perception towards more saturated colours when delay length was increased, because saturated colours are more behaviourally relevant for communicating about objects.

In summary, using a delayed colour estimation task across hue and saturation, a 2D model of the geometry of perceptual colour space was proposed. Previous research has found evidence for categorical biases in colour perception across several category boundaries, but this is the first time that these biases have been mapped out across hue and saturation. The evidence supports a prior which biases perception and memory towards saturated colour categories. In particular, responses were biased away from the achromatic category foci (grey) and towards five chromatic category foci (green, orange, red, pink and blue). The nature of this prior was outlined, but its origin was merely speculated. This is an important avenue to investigate in order to fully understand the forces which influence colour perception.

4 The amount of information provided by colour about object identity using a kernel density estimation model.

Colours are useful for object recognition (Bramão et al., 2011), therefore categorical biases in colour perception may originate from the distribution of colours across objects. Here, object colours were sampled from ImageNet, and kernel density estimation was used to estimate the probability distributions of colour for a given object. Bayes' rule was applied to calculate the probability of an object given a colour, and variation in probability distributions was measured using information theory. It was predicted that colour would provide 2.58 bits of information about objects on average, and the information geometry of object colours would reflect the distribution of BCTs across the space, such that category foci are the most informative about object identity. Colour only provided 1.32 bits of information about objects. Information geometry did not vary in a similar structure to BCTs, but there was an effect of saturation: high saturation and zero saturation colours were more informative about objects than low saturation colours, reflecting the average probability of colours in the environment. As such, the model successfully grouped objects into chromatic and achromatic colour categories.

4.1 Introduction

In the previous chapter, categorical biases in perceptual colour space across hue and saturation dimensions were identified. Previous research has found evidence for categorical biases in colour perception across several category boundaries, but this is the first time that these biases have been mapped out across the entire visible range of colours. The patterns of categorical biases were stable across participants and delay conditions.

Although categorical biases in colour perception have been established, the origin of these colour categories are hotly debated (Regier & Kay, 2009; Siuda-Krzywicka et al., 2019; Witzel, 2019). The universality literature traditionally argued that these biases originate from biological constraints on colour perception, which are hard-wired into the human visual system (Heider & Olivier, 1972; Kay & McDaniel, 1978; Saunders & van Brakel, 1997). Kay and McDaniel (1978) proposed that the 11 BCTs observed universally (Berlin & Kay, 1991) are a combination of the unique hues (red-green, blue-yellow and black-white) excited by colour opponent processes in the retinogeniculate pathways.

These unique hues were argued to be perceptually salient, with enhanced memory and learning for them regardless of whether colour terms existed for these colours in the participant's language (Heider & Olivier, 1972).

On the other hand, relativists argue that categorical colour perception is linguistically determined, whereby colour categories in language and culture shape colour perception (Jameson, 2005; Roberson et al., 2000; Whorf, 1956). Evidence for linguistic relativity comes from significant variation in the size and location of colour lexicons across languages, which influences colour perception (Davidoff et al., 1999). Participants show faster and more accurate discrimination for colours across category boundaries than colours within the same category, but only for colour categories in their own language (Hu et al., 2014; Kay & Kempton, 1984; Özgen & Davies, 2002; Roberson & Davidoff, 2000).

It is clear that there is a diverse range of colour lexicons across the world which do explain variation in colour categorisation. Despite this, there are also notable patterns of similarity between languages. Berlin and Kay (1991) identified that of 2048 possible combinations of 11 BCTs, only 22 exist. Language has been used to explain cross-cultural variation in colour categorisation, but it's possible that constraints on language could also explain cross-cultural similarities. Each culture will acquire linguistic colour terms which are most useful and relevant for their functional needs. These functional needs may be shared across cultures, resulting in cross-cultural similarities in colour categorisation and perception.

All languages universally have the same problem: how should a continuous colour space be categorised to communicate about the colours in the environment most efficiently? Humans are able to perceive over two million distinguishable colours (Linhares et al., 2008; Pointer & Attridge, 1998), yet most languages have a maximum of 12 BCTs (Berlin & Kay, 1991). As such, there are two cultural forces which constrain colour lexicons: communication efficiency and the statistics of colour in the environment. Communication efficiency defines the balance between a language having enough linguistic terms which precisely communicate important information, whilst not having too many unnecessary terms which complicate the language and increase cognitive demand. An efficient colour lexicon should be organised such that there are category boundaries between colours which are useful to discriminate between. The colours which are behaviourally relevant for categorisation are determined by the statistics of colours in the environment, i.e. how colours vary across objects. For example, since a banana varies in colour from green to yellow to brown, depending on its ripeness, these three colour categories are a useful and efficient way to communicate the ripeness of a banana.

There may be some variation in these forces, where some colours are more behaviourally relevant in certain habitats. For example, the arctic is covered in snow and ice making white colours more prevalent and behaviourally relevant to residents of that region. This leads to more colour terms for white in their local language (Fortescue, 2016). However, in general there are a lot of statistical similarities in colours across the world. Webster et al. (2007) took images of natural outdoor scenes in the Western Ghats (a subtropical mountain range in India) and Sierra Nevada (dry/snowy mountains in the USA) during different weather seasons, and observed similar seasonal variation in colour statistics of the environment for both locations. Humans may adapt to the distribution of colours in the environment, which could explain universal patterns in colour categorisation.

The goal of this study is to model the nature of communication efficiency and the colour statistics of the environment, which constrain colour categorisation. An optimal language will increase communication efficiency for the most behaviourally relevant colours. The colours which are most behaviourally relevant depends on the function of human colour vision.

Many studies have found evidence that coloured objects are easier to recognise than black and white objects (Davidoff & Ostergaard, 1988; Laws & Hunter, 2006; Price & Humphreys, 1989; Tanaka & Presnell, 1999; Wurm et al., 1993). In a meta-analysis of 35 studies which compared object recognition for regular versus greyscale images of objects, Bramão et al. (2011) found strong evidence that colour information improves object recognition. In addition, Ennis et al. (2018) measured the distribution of colour pixels for fruits and vegetables in CIELAB colour space, and found that all the pixels for a single object fall within a very narrow distribution of hues. They therefore concluded that hue can define a fruit or vegetable. This research suggests that colours are useful to help humans identify objects, therefore the colours which are most behaviourally relevant depends on the colours of objects in the environment.

For the task of object recognition, an optimal representation of colour should maximise differences between object classes, while minimising differences within classes. So, it is possible that category boundaries between BCTs are useful for discriminating between different objects in the environment, and thus these colour categories are observed universally because they are useful to inform humans about objects.

Here, a Bayesian model of colour perception is proposed, where biases observed in human colour perception are due to a non-uniform prior over colours. Sensory signals contain uncertainty, due to internal or external noise (e.g. cognitive load or lighting conditions; Olkkonen et al., 2014). The likelihood of a signal varies as a function of noise: under noisier conditions, the likelihood of a signal has more variation, so there is more uncertainty. To reduce uncertainty, these noisy sensory signals are combined with prior knowledge about the world. Priors are an internal representation of environmental statistics, which is learned through exposure to the environment. If there are no patterns in the statistics of the environment, then the prior probability will be uniform. However, if there are statistical patterns in the environment, then a non-uniform prior will be learnt. The prior is combined with the likelihood of a signal to compute the posterior probability. As such, the prior facilitates perception of noisy signals by directing observers away from unlikely conclusions and towards likely conclusions, based on experience with the world. Since colours are used to improve object recognition, a non-uniform prior over colours may originate from the distribution of colours in the environment. It is proposed that the prior directs noisy sensory signals towards colours which are more informative about objects and away from less informative colours, resulting in biases in colour perception. As such, category foci are expected to occur around informative colours, and category boundaries are expected around uninformative colours.

To identify the form of this prior, kernel density estimation was used to model the association between objects and their surface colour. The model is fitted to a dataset containing object identities and their associated colours, and returns the probability distribution of colours given the object identity. The aim of this study is to identify how much information is provided by colour about objects. It was predicted that the information geometry of the colour space would be closely related to the structure of perceptual colour representation, such that colour perception is biased towards category foci because focal colours are most informative about object identity.

4.2 The dataset

The images used were sampled from Imagenet (Russakovsky et al., 2015). The open source dataset consists of over a million images of objects and scenes organised into 1000 object classes. Amongst many others, object classes range across different types of vehicles, artefacts, foods, buildings, plants and animals (including 120 dog breeds for testing fine-grained discrimination). The dataset was collected by querying the object classes on internet search engines and automatically downloading them. Therefore there is large variation in lighting conditions, viewing angle, object size and background. In addition, images collected from unknown camera outputs which may not have been calibrated, or may have different calibration settings, will introduce some unknown variation; images taken of the same scene using different cameras do vary (Stevens et al., 2007). It was assumed that since the dataset is enormous, any device dependent variation is justified by the large number of devices used to collect the data, and will contribute to the normal distribution of noise across images. Therefore the object colours extracted and converted into perceptual chromaticity spaces are treated as an approximation of the object surface chromaticity.

Human annotators recruited through Amazon Mechanical Turk were used to accurately annotate objects in images and identify their bounding boxes (Russakovsky et al., 2015). Between 2 and 5 participants were required to verify each image label.

For the current dataset, 118 of the dog classes were collapsed down to one generic dog class with 600 randomly sampled dog images, to ensure that all classes were of similar hierarchical level. In total, 430930 images were used, organised into 883 object classes, resulting in 488 images per class on average. The images were cropped to a bounding box to remove the background. The cropped image was then split into a 3x3 grid, and a random pixel was selected from the central section of each image (an alternative sampling method was also tested – see Appendix C). The RGB value of this pixel was used to represent a colour within the object. As such, there was a single RGB value for each image in the database, along with its object classification.

The resulting RGB values were converted into CIE1931 xyY chromaticity coordinates using the CIE RGB to XYZ transformation matrix (Lindbloom, 2017) and XYZ to xyY transformation (Equation 4.1). The Y value, representing luminance, was removed, and only xy chromaticity coordinates were analysed, to represent hue and saturation dimensions of colour space. As luminance information was removed, there is no distinction in the data between colours which only occur at specific luminance levels, such as black/white/grey, and yellow/brown.

Luminance information was removed from the data because luminance is heavily influenced by illumination (Kingdom, 2011). Humans have lightness constancy, which achieves a robust sensation of surface reflectance despite significant variation in lighting conditions. However, lightness constancy requires cues about illumination to resolve ambiguity about surface reflectance. By sampling a single pixel from the image, any cues to illumination have been removed, therefore luminance is not a reliable signal of surface reflectance. It is possible to model information about surface reflectance (Baddeley & Attewell, 2009), but it’s not within the realm of this model: the focus of the current model is on chromaticity.

Equation 4.1 *Transformation functions from xyz to xyY.*

$$x = \frac{X}{X + Y + Z} \qquad y = \frac{Y}{X + Y + Z} \qquad Y = Y$$

4.3 Kernel density estimation

Kernel density estimation (KDE) was used to fit the data, as it can predict a smooth distribution of probabilities across multiple variables which vary non-parametrically. These predictions can be used to make inferences about the data, based on a finite data sample. It is used here to predict the probability of a colour, given an object. This can be understood as a two-dimensional smoothed histogram. If there are more colours grouped locally in the 2D colour space, the estimated likelihood of the object is higher, as the probability of a colour at that location for the object increases.

The model fits the data by applying a kernel smoothing function to the input data to generate an estimate. Figure 4.1 demonstrates an example of this fit for a random, bimodal distribution varying in one dimension. Smoothing is implemented by averaging neighbouring datapoints, thus blurring any sharp edges in the data. The weight given to neighbouring points is determined by the type of kernel and the bandwidth. A Gaussian kernel was used, where a Gaussian curve is centred on a datapoint, and neighbouring points are divided by the area under the Gaussian curve.

The bandwidth is the parameter that determines the amount of smoothing, thus controlling the trade-off between signal and noise in the model output. The bandwidth is essentially the standard deviation of the Gaussian curve used to implement smoothing. A large bandwidth leads to oversmoothing, such that more weight is given to points which are further away, therefore the predicted distribution contains high bias and important signals are lost. A small bandwidth leads to undersmoothing, where very few neighbouring points are given weight, so the predicted distribution contains a lot of noise. Bandwidth selection is therefore important to achieve the optimum signal-noise ratio from model predictions.

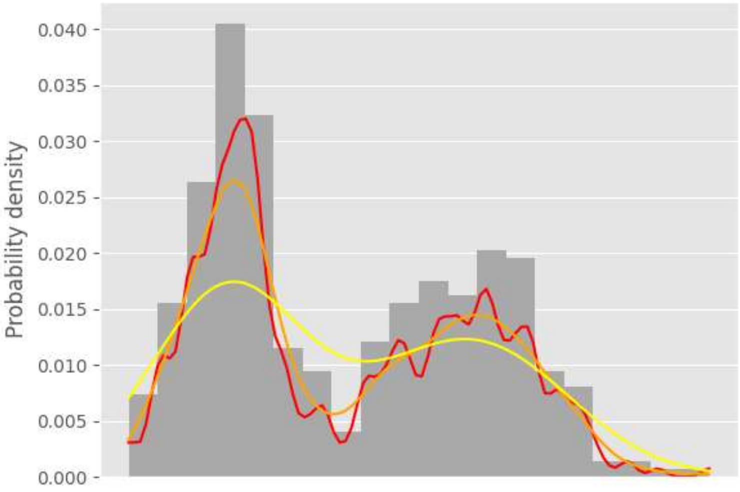


Figure 4.1 *An example of 3 KDE models fit to a random bimodal probability density function varying in one dimension. The KDE models are fit with a bandwidth of 1 (red), 3 (orange) and 8 (yellow).*

Methods

For each object class, KDE is implemented using the statsmodels Python library (Seabold & Perktold, 2010) to fit the probability distribution of colours for a given object.

To select the bandwidth, Scott's rule of thumb was implemented (Scott, 2012). Scott's rule is efficient and robust, and is appropriate for multivariate, continuous, nonparametric data. It is important to select a suitable bandwidth for each individual object, since the distribution of colours varies between different objects. Some objects may have very little variation in colour (e.g. black bear), therefore a large bandwidth will be unlikely to introduce noise into the signal. In contrast, multicoloured objects (e.g. toy car) will benefit from a smaller bandwidth to ensure that the multimodal signal is not lost.

Once the model has been fitted to the training data, it is evaluated across the RGB triangle in CIE1931 chromaticity space (Smith & Guild, 1931). The model generates estimates for the probability of each colour given the object P(colour|object) for each of the 883 object classes. Figure 4.2 shows the distribution of observed data for three objects with the distribution of model predictions for the associated objects. The model predictions vary smoothly and reflect the distribution of the observed data, demonstrating that the model provides a good fit.

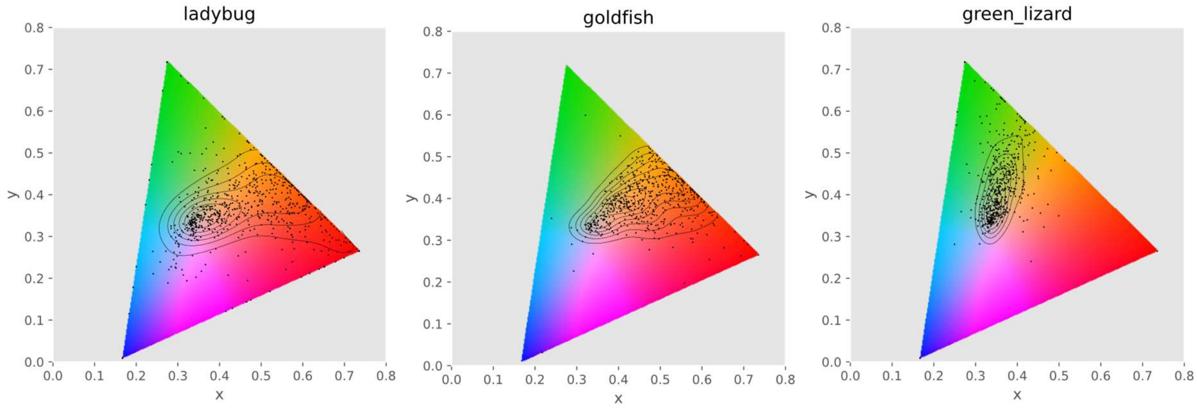


Figure 4.2 *Probability distribution of colours which predict an object class P(colour|object) for object classes: "Ladybird", "Goldfish" and "Green Lizard".Scatterplots show the observed data for that object, and contour plots show model estimates.*

The goal of this study is to quantify how much information colour provides about object identity. Bayes' rule (Equation 4.2) was used to calculate the probability of an object given the object colour P(object|colour).

Equation 4.2 *Bayes' rule for P(object|colour).*

$$P(object|colour) = \frac{P(colour|object) \cdot P(object)}{P(colour)}$$

Information theory was then applied to quantify the amount of information that colour provides about objects by measuring variation in probability distributions. Variation in probability distributions was measured using entropy, which is a method of characterising uncertainty in a probability distribution. Increasing the number of possible outcomes or the uniformity of the probability distribution will make it more difficult to predict the outcome, therefore increasing the uncertainty, and thus the entropy. The maximum entropy for a uniform probability distribution with *n* outcomes, is log2(n). For example, given the task of predicting the outcome of a fair, 8-sided die, there is a 1/8 chance of getting the outcome, and the entropy is log2(8)=3bits. Consequently, predicting the outcome of a fair, 16-sided die will increase uncertainty, and entropy is log2(16)=4bits. Predicting the outcome of a die biased toward landing on even numbers will reduce uncertainty, as the underlying probability distribution is no longer uniform, making the task of estimation easier and reducing entropy. The entropy of a non-uniform probability distribution is calculated using Equation 4.3.

Equation 4.3 *A random variable X with a probability distribution p(X) has an average entropy H(X).*

$$p(X) = \{p(x_1), \ldots, p(x_m)\}$$

$$H(X) = -\sum_{i=1}^{m} p(x_i) \log_2 p(x_i)$$

If the probability distribution associated with a given colour is uniform, then all objects are equally as likely, therefore the estimate would be a random guess. However, if the information provided to the model in the training data was useful for object identification, then the model will have learnt the colour properties of each object. Therefore, for a given colour, the model will predict some objects to be more probable than others, resulting in a non-uniform probability distribution.

It was predicted that the model would organise objects into colour categories, such that there is less uncertainty about object identity at category foci and more uncertainty at category boundaries. There are 11 BCTs in total, however six colour categories were identified in the perceptual chromaticity space in Chapter 3 (red, green, blue, orange, pink and grey). Therefore, it was predicted that the model would organise objects into six colour categories. The entropy of a uniform distribution of six colour categories is log2(6)=2.58 bits. If the model organises objects into six equally probable colour categories, then colour should provide 2.58 bits of information about object identity.

4.5 Results

The mutual information between object classes and colours was measured across the colour space using information theory. Mutual information quantifies the amount of information obtained about one random variable, by observing another random variable.

Table 4.1 *Mutual information between object classes and colours for a KDE model.*

	KDE Model
Uncertainty about colour, H(colour)	11.60 bits
Uncertainty about colour given object class, H(colour|object)	9.00 bits
Average amount of information provided by objects about colour, I(colour|object)	11.60-9.00 = 2.60 bits
Uncertainty about object class, H(object)	9.79 bits
Uncertainty about object class given colour, H(object|colour)	8.47 bits
Average amount of information provided by colour about objects, I(object,colour)	9.79-8.47 = 1.32 bits
Average number of colour categories used to discriminate between objects	$2^{1.32}$ = 2.50

4.5.1 What information do objects provide about colour?

How much uncertainty is there about colour?

Given no information about an object, estimating the object colour from all possible colours is a random guess. Uncertainty about colour is high, as all colours in the colour space are equally as likely. The entropy of the list of average probabilities of each colour H(colour) was 11.60 bits (Table 4.1).

How much uncertainty is there about colour, given the object identity?

To measure uncertainty about colour given the object identity, the entropy of the probability distribution of colours was calculated for each individual object H(colour|object). If uncertainty varies across different objects then it suggests that some objects are more informative about colour than others. The entropy was averaged across all objects, to measure the average uncertainty about colour once object identity is known, which was 9.00 bits (Table 4.1). Figure 4.3 shows the probability distributions of colour for a given object P(colour|object). Colour is more variable for some objects than others, but the model has successfully learnt representations of object colour. Object identity is informative about object colour, as the probability distributions are not uniform.

How much information do objects provide about colour?

The difference between uncertainty about colour given no information about the object, and uncertainty once object identity is known, is the amount of information provided by objects about colour (Equation 4.4). Figure 4.4a shows the distribution of information values across different objects. Information is normally distributed about a mean of 2.60 bits. Therefore, on average, objects provide 2.60 bits of information about colour (Table 4.1).

Equation 4.4 *Amount of information provided by colour about objects.*

$$I(object, colour) = H(object) - H(object|colour)$$

Some objects provided more information about colour than others (Figure 4.4b, from a minimum of 0.18 bits for “Balloon” up to a maximum of 6.66 bits for “Geyser”. This mutual information can be thought of as a measure of colour diagnosticity, where colour diagnosticity is the degree to which a colour is associated with an object, so the colour of a highly colour diagnostic object is more accurately predicted. A low colour diagnostic object will have variable colour properties, therefore it is more difficult to predict the object’s colour from just the object identity.

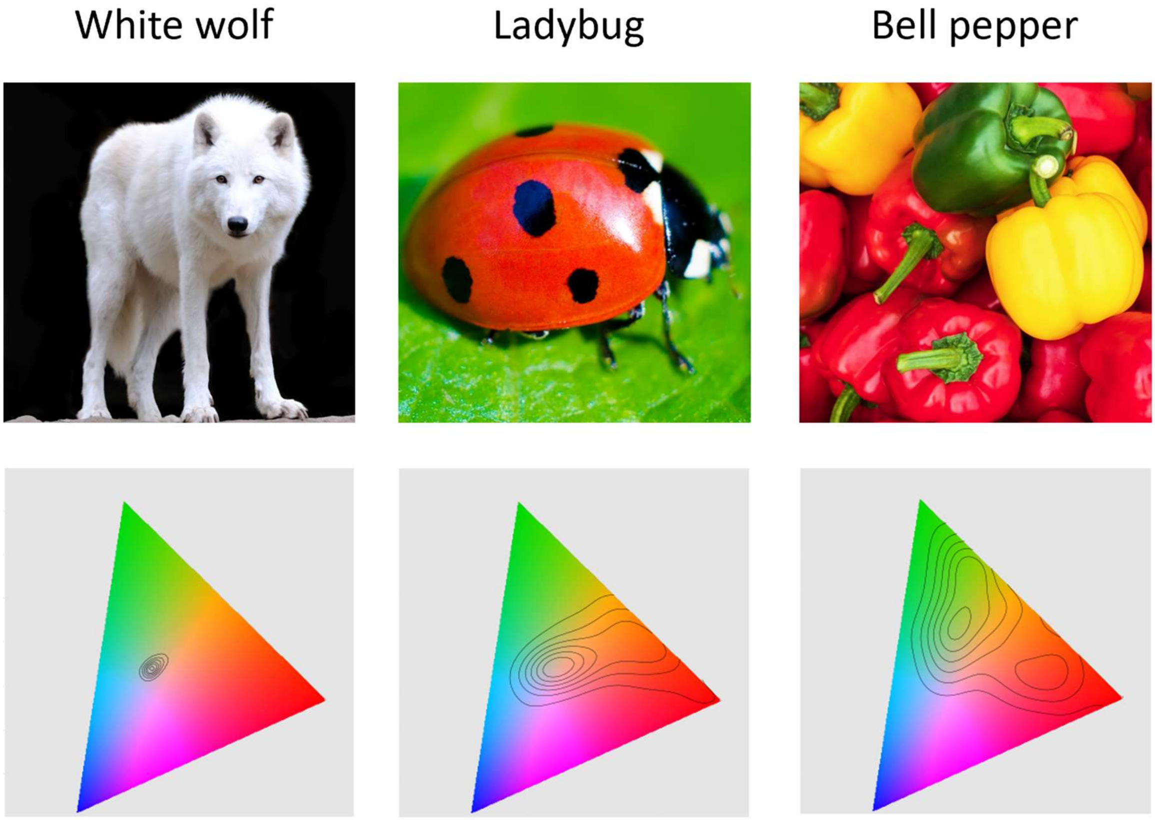
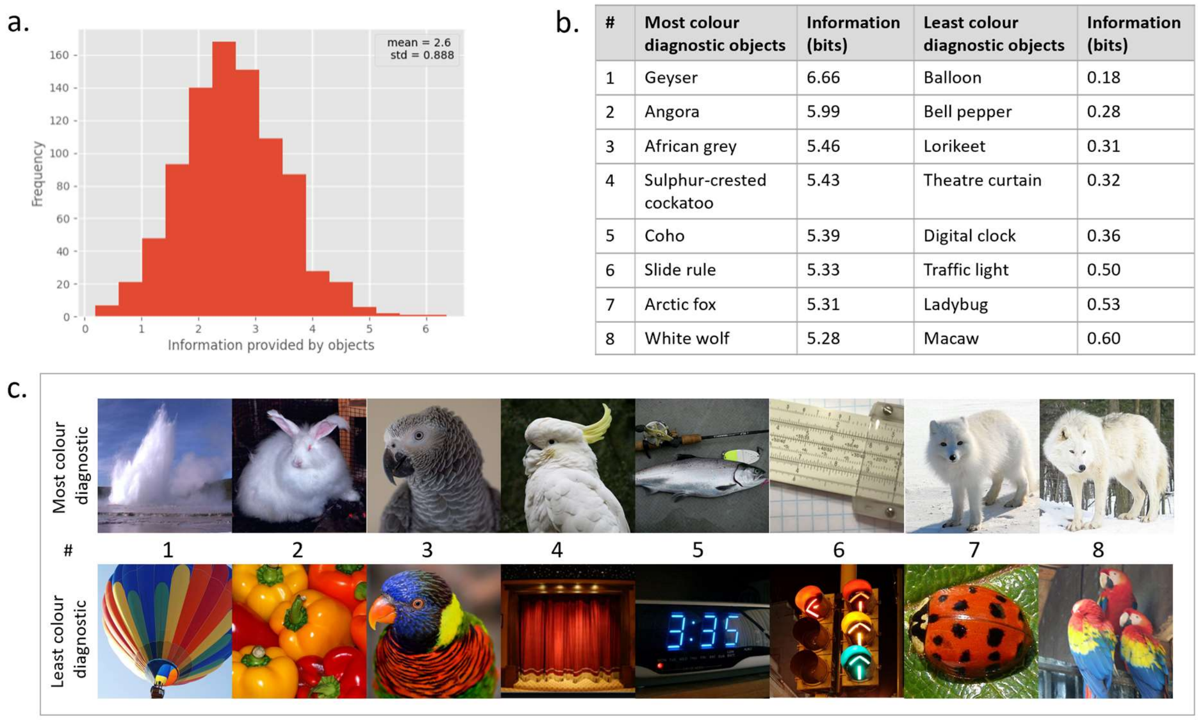


Figure 4.3 *Model estimates for colours given the object identity P(colour|object) for object classes “White wolf”, “Ladybug” and “Bell pepper”. Images show examples of the objects from ImageNet, and contour plots show the probability distribution of colours across CIE1931 chromaticity space given the object identity.*



#	Most colour diagnostic objects	Information (bits)	Least colour diagnostic objects	Information (bits)
1	Geyser	6.66	Balloon	0.18
2	Angora	5.99	Bell pepper	0.28
3	African grey	5.46	Lorikeet	0.31
4	Sulphur-crested cockatoo	5.43	Theatre curtain	0.32
5	Coho	5.39	Digital clock	0.36
6	Slide rule	5.33	Traffic light	0.50
7	Arctic fox	5.31	Ladybug	0.53
8	White wolf	5.28	Macaw	0.60

Figure 4.4 *Information provided by colour about objects for a KDE model. (a) Probability distribution of information values across objects. (b) Table showing the most and least colour diagnostic objects with their associated information value. (c) Images sampled from ImageNet from the most and least colour diagnostic objects. The number corresponds with the number in the first column in (b).*

4.5.2 What information do colours provide about objects?

The previous section confirmed that these models provide a suitable fit to the data, and also produced quantifiable outcomes for colour-concept associations; there is an association between objects and their surface chromaticity. The model produced representations of object colour and used these representations to make future predictions.

The next step in this analysis was to assess whether associations between objects and their colour properties can be used to improve object recognition. This was achieved by measuring how the probability of objects vary across colour space.

How much uncertainty is there about objects?

Given a task of identifying one object out of 883 equally possible objects, with no information about that object, there is a 1/883 chance of correctly estimating the object. The entropy of a uniform distribution with 883 possible outcomes is log2(883) = 9.79 bits. The entropy of the list of average probabilities of each object *H(*object) was 9.79, showing that all objects are equally as likely (Table 4.1). This is the maximum uncertainty about objects.

How much uncertainty is there about objects, given the colour?

Uncertainty about object identity given the object colour was then measured by calculating the entropy of the probability distribution for each individual colour $H(\text{object}|\text{colour})$. The entropy was averaged across all colours to estimate average uncertainty about objects given object colour, which was 8.47bits (Table 4.1).

How much information does colour provide about objects?

The difference between uncertainty about object identity given no information about the object $H(\text{object})$, and uncertainty once colour is known $H(\text{object}|\text{colour})$, is the amount of information provided by colour about objects (Equation 4.4). The average amount of information colours provide about objects was 1.32bits of information (Table 4.1). This totals to $2^{1.32}=2.50$ colour categories, whereas it was predicted that the model would organise objects into six colour categories.

Amount of information provided by colour about objects was then plotted on a contour plot over the CIE1931 chromaticity space (Figure 4.5b). The largest effect on information was saturation, where information about objects increases with increasing saturation.

There do appear to be some interesting contours around unsaturated colours. However, as the information values are not very normally distributed (Figure 4.5a), it is difficult to view the information geometry on the contour plot. Therefore, a log transformation was applied to the information values (Figure 4.5c-d) which allowed the geometry to be more visible. From the log transformed plots, it is clear that while less saturated colours are less informative, colours with zero saturation (black, white, grey) are more informative than colours with low saturation.

In addition, there seems to be an effect of the Planckian Locus. The Planckian Locus is a chromaticity spectrum determined by the temperature of a black body which emits light as a result of being heated (Koenderink, 2010). The sun is an example of an incandescent black body, therefore natural daylight closely parallels the Planckian Locus; from morning blue light at sunrise, through to white, yellow, and finally orangey-red during sunset[1]. Low saturation colours that lie on this Locus are more informative than their equally as saturated neighbours (Figure 4.5e). However, for high saturation, colours which are further from the axis are more informative. This trend seems to reflect the distribution of the average probability estimate of colours within the model (Figure 4.5f). Colours along the Planckian Locus have a higher probability.

---

[1] The daylight locus is slightly different to the Planckian Locus due to air particles which refract light on Earth (Rayleigh scattering), but the difference is small and both loci follow a similar trajectory (Hernández-Andrés et al., 2001). As such, the term Planckian Locus is used to refer to the daylight locus and Planckian Locus interchangeably.

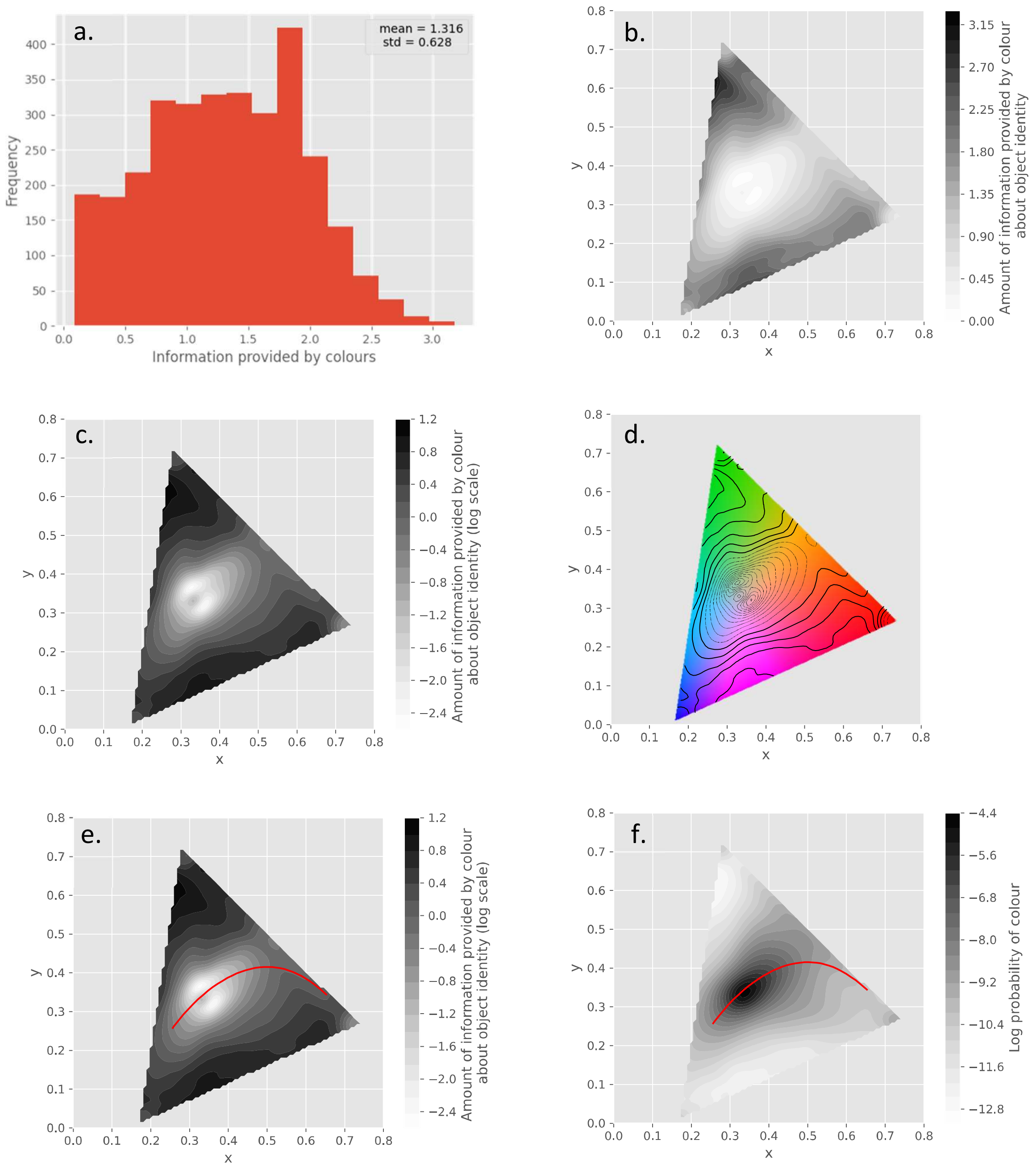


Figure 4.5 *Information provided by colour about objects for a KDE model. (a) Histogram showing the distribution of information values across colours. (b) Amount of information provided by colour about objects across CIE1931 chromaticity space. (c) Figure 5b transformed onto a log scale. (d) Identical contour plot to Figure 5c with the colour space displayed underneath. (d) Contour plot showing the log amount of information provided by colour about object identity with the Planckian Locus plotted in red. (e) Contour plot showing the log average probability of each colour within the model, with the Planckian Locus plotted in red.*

4.6 Discussion

Since six BCTs were identified in the chromaticity space for high-level perception, it was predicted that colour would provide log2(6)=2.58 bits of information about object identity. However, in this model, colour provided only 1.32bits of information. In addition, it was predicted that the information geometry of colour space would reflect the foci and boundaries of colour categories, where category foci should be most informative about objects. The model did not organise colours into colour categories. However there was an effect of saturation where, in general, information increased with increasing saturation. This saturation effect interacted with the Planckian Locus, where colours with low saturation were more informative when close to the Locus, but colours with high saturation were less informative when close to the Locus.

A peak in information suggests that these colours are more useful than their neighbouring colours for identifying objects, as the predicted probabilities are less uniformly distributed. In the present model, high saturation colours are more useful for identifying objects than low saturation colours. This is due to low saturation colours being much more common in the dataset than high saturation colours (Figure 4.5f), thus it is more difficult to identify an object given a low saturation colour, because there is a larger number of possible objects. For example, learning that an unknown object is a beige colour is not very informative about the object identity, because many objects are beige coloured. Whereas, learning that an object is red is informative, because there are fewer red objects, so the colour red significantly reduces uncertainty about object identity.

Not only are low saturation colours much more common than saturated colours, but also colours which lie along the Planckian Locus are more likely. Therefore, due to their proximity with the Planckian Locus, saturated blue and yellow are also less informative than saturated red, pink and green.

Using different datasets and methods, other researchers have also observed that the distribution of colours in the environment tends towards relatively unsaturated colours, with most variation in chromaticity being along the Planckian Locus. Analysing the colour statistics of a database of over 20,000 images, Rosenthal et al. (2018) found that both objects and backgrounds (natural and artificial) tended to be of a low saturation. Yendrikhovskij (2001) used a clustering algorithm to categorise all of the pixels from 630 natural images, and found that low saturation colours were more common than high saturation colours. Webster et al. (2007) compared images of natural outdoor scenes in the Western Ghats (India) and Nevada (USA) during different seasons. They found that the colour distribution of sky in all environments and the earth in dry environments tend to vary along the daylight locus. Therefore the distribution of colours observed in the present study is not just a property of the objects in the ImageNet dataset used, but also reflects the distribution of colours in the environment, as observed in multiple studies using different datasets and methods.

Human colour vision adapts to the distribution of colours in the environment to support perception of the most behaviourally relevant stimuli (Shepard, 1992; Webster, 1996). Webster and Mollon (1997) observed that the average chromaticity of natural scenes tends to vary along the Planckian Locus in natural environments, and that participants adapted to the colour distribution of these environments by displaying selective changes in colour appearance towards the adapted colour distribution in a colour matching task. Bosten et al. (2015) found that achromatic adjustments towards unique white varied most strongly along the Planckian Locus, which suggests that the range of colours perceived as white is calibrated to the distribution of colours in the environment. Participants have also shown improved discrimination and less variability in colour matching for colours along the Planckian Locus (Danilova & Mollon, 2010; Panorgias et al., 2012). In addition, fMRI responses to V1 in macaques were stronger for colours along the Planckian Locus, despite no differences in responses in the LGN (Lafer-Sousa et al., 2012).

The findings reported here lend further support to the idea that colour vision is adapted to the colour distribution of objects in the environment, by directing colour perception towards the most informative colours in the form of prior expectations. In Chapter 3, it was observed that responses in a colour estimation task were biased towards more saturated colours (Figure 3.9). The current model can explain these biases, because saturated colours are less common than unsaturated colours, and therefore more informative about objects. As such, it would be more useful for object recognition to perceive saturated colours, so prior expectations direct perception towards saturated colours.

Although the distribution of colours in the environment can explain the majority of the information geometry observed in this model, there is one effect that differs. A peak in information was observed for colours with zero saturation (i.e. white, grey and black); however there was no difference in the likelihood of low saturation and zero saturation colours. One possible explanation for this observation is that the objects which provide the most information about colour were achromatic (Figure 4.4c). Objects which are white or grey and under a single illuminant exhibit no variation in chromaticity, because the chromaticity of the illuminant and the object is the same (J. Mollon, 2006). There is also no luminance information in the CIE1931 chromaticity space, therefore achromatic colours occupy a very small region in this space. As a result, achromatic colours are highly informative about achromatic objects. Thus the model learnt to discriminate between objects which are chromatic and achromatic, which explains the peak in information for achromatic colours.

This highlights an issue within previous literature on the role of colour information on object recognition. The model identified objects which are colour diagnostic *and* achromatic. Evidence that colour information improves object recognition normally involves comparison of object recognition for regular versus greyscale images of objects, and object recognition is usually better for objects with colour information (Bramão et al., 2011). However, there isn't any acknowledgement of the existence of achromatic objects (both natural and artificial), whose achromatic surface is also informative about object identity. It has been proven that chromatic colours are useful for object recognition by removing chromatic information from images, but the usefulness of achromatic colours for object recognition has been neglected.

Interestingly, the objects with lowest colour diagnosticity in this model tend to have high saturation (Figure 4.4c), despite the finding that high saturation is more informative about object identity. Many of the objects are either multicoloured ("Lorikeet", "Macaw") or have variable colours across objects ("Balloon", "Bell pepper"). Therefore the colour properties of these objects are widely distributed across the space. The more colours which are probable for a given object, the more uniform the probability distribution is, therefore there is greater uncertainty about colour.

This highlights a limitation of a KDE model for measuring colour-concept associations. The model is not able to differentiate between multicoloured objects and objects with variable colour properties, e.g. ladybirds are red and black, whereas bell peppers are red or orange or yellow or green, but both are low in colour diagnosticity. Ladybirds are actually high colour diagnostic objects, since they are reliably coloured, but the model is unable to recognise this.

According to the global assignment hypothesis, colour diagnosticity is not simply a factor of the colours that are most strongly associated with a given object, but also whether the colour of the object is a distinguishing feature that differentiates that object from other objects (Rathore et al., 2020; Schloss et al., 2018; Tanaka & Presnell, 1999). Ladybirds are also colour diagnostic because there aren't many objects which have a similar colour distribution, therefore its colour properties distinguish a ladybird from other object classes, thus providing useful information for object recognition.

Tanaka and Presnell (1999) measured colour diagnosticity as a combination of typicality judgements and feature listing. In the typicality judgement task, participants named the most typical colour associated with an object. The feature listing task asked participants to list the top three perceptual features associated with an object. If a colour is mentioned as a top perceptual feature for an object, then it suggests that colour is a distinguishing feature for the object. For the 48 objects in their study, typicality judgements and feature listing were only weakly correlated (r=.543). This suggests that the information provided by objects about colour is not a direct measure of colour diagnosticity. In fact, humans often associate objects with colours that are not properties of the object. For example, blueberries are not very blue, but they're associated with blue because they're the bluest berry (Rathore et al., 2020). It would be valuable to test whether a model that estimates the probability of objects for a given colour is able to learn the colours that distinguish objects from each other, and thus provide an alternative measure of colour diagnosticity.

Although KDE is useful for estimating the colours that are most strongly associated with a given object, it is not very effective for identifying the distinctive colour properties of objects. When used to compare the probability of objects for a given colour, the model primarily fitted to the average probability of colours across the space, where unsaturated colours and colours along the Planckian Locus are most common. This resulting in an effect of saturation and the Planckian Locus on information geometry, but no categorical biases across hues. Therefore, the results of this study can explain biases in colour perception towards saturated colours, but the origin of biases towards the foci of BCTs is left to be determined. A more intelligent method of modelling the distribution of object colours may be more successful at learning complex patterns in chromatic colours across objects.

5 The amount of information provided by colour about object identity using a deep neural network.

A deep neural network was trained to identify objects using a single pixel sampled from object images, to test whether complex patterns of information about objects can be identified across the colour space. The model generated a probability distribution across objects for a given colour. Information theoretic methods were used to analyse how predictions varied across the colour space, and how informative colours are about objects. It was predicted that the model would organise objects into colour categories and, on average, that colour would provide the same amount of information as the entropy of six BCTs. The model organised information provided by colour about objects into five distinct colour categories: red, green, blue, pink and grey, where chromatic colours were most informative about objects, and achromatic colours were least informative about objects. On average colour did not provide the same amount of information as six colour terms, because the colours were not uniformly distributed across the chromaticity space. However, it was observed that distortions to the geometry induced by the object recognition task reflected the non-uniform prior in colour perception (measured in Chapter 3), such that colours which were most informative about objects were the colours towards which colour perception was biased. In addition, the colours which were least informative about objects were those which perception was biased away from. These findings support the theory that the colour statistics of objects in the environment may form the basis for a non-uniform prior which directs perceptual processes towards the most informative colours.

5.1 Introduction

In the previous chapter, kernel density estimation was used to model the colours most strongly associated with an object, and measure how much information colours provide about those objects. The model successfully estimated the probability of colours given an object, and thus produced representations of object colours. However, information provided by colour about objects was mostly determined by the average probability of colours in the dataset, and complex patterns of information across objects were not recognised. In particular, saturated colours are unlikely in the dataset compared with unsaturated colours, and so category boundaries between saturated colours were not detected in the information geometry.

In this chapter, a deep neural network is used to test whether more complex patterns of information can be learnt across the colour space. Deep neural networks don't make any assumptions about the distribution of the input data, and their multiple hidden layers allow them to learn and detect complex, non-linear relationships in the data. As such, they are very good at pattern recognition and deliver very efficient results, thus making them a good candidate for modelling the relationship between objects and their colour properties. In addition, a neural network estimates the probability of objects for a given colour, and thus it may be able to learn the colours which differentiate objects from each other.

5.2 Deep neural network

Deep neural networks are machine learning algorithms that can achieve pattern recognition with very high accuracy. They receive a list of input signals and transform them into output signals through multiple hidden layers of computational units. Each layer of computational units learns to detect recurrences of values. Multiple pathways between these layers enable the network to recognise complex patterns in the input data, which can be used to predict the outputs of new, unseen data.

In a standard neural network for object recognition, the model is given information about each object in the form of labelled images (Bishop & Bishop, 1995). The model detects patterns in object properties, and learns to identify objects in images. Given a new image, the model outputs a vector of probabilities, which sum to 1. These indicate the probability of each object for the given image, which can be used to make an estimate of the object identity.

An image contains a lot of information about the object, including colour, shape, texture and context. The aim of the work reported here is to identify how much information is provided by colour about objects. To investigate this, a model was trained to identify objects using a single colour representing the object colour, and the associated object's label. This greatly reduces the amount of information provided to the model about the object, and creates a very difficult task: one that even a human would not be capable of accomplishing accurately. If colour provides no information, then the model will perform at chance. The model estimates will be uniformly distributed, with maximum uncertainty. However, if colour is useful for object identification, then the model will perform above chance, and uncertainty will be reduced, suggesting that some information about object colour was learnt.

Once the model was trained to recognise objects using colours, the model was tested to measure how information about objects varies across a continuous colour space. It was predicted that the structure of this space would be closely related to the structure of perceptual colour space, and to the structure of BCTs, such that colour perception is biased towards category foci because focal colours are most informative about object identity. In addition, it was predicted that the model would organise objects into the six BCTs in the colour space, and therefore that colours would provide log2(6)=2.58 bits of information about object identity.

5.3 Methods

A deep neural network was trained to classify objects using the dataset sampled from ImageNet outlined in Section 4.2 (an alternative dataset was also tested – see Appendix C). Several networks with different architectures were tested, and consistent results were observed; therefore one model architecture is described for clarity. The deep neural network models were built using Keras (Chollet, 2015), a Python API for deep learning running on top of the neural network library TensorFlow (Abadi et al., 2016).

The architecture of the network consisted of 4-layers, with 3 fully connected layers (512 units per layer with rectified linear unit activation) and one classification layer (with softmax activation). The weights were initialised using Glorot uniform distribution (Glorot & Bengio, 2010). Batch normalisation was also implemented after the first 2 layers to normalise the inputs into a layer, stabilise learning and reduce the number of epochs required to train the network. Batch size was set to 128 and an Adam optimiser (Kingma & Ba, 2017) was used. The network is trained to classify objects. The training data had 2 features (x and y chromaticity coordinates), 430930 samples and 883 classes of objects.

The model was given very little information about the object, so there was a lot of variance in the output, particularly among predictions for unlikely objects. Therefore, an ensemble of 30 separate models was trained, and their predictions were averaged to reduce noise, produce much smoother results, and improve performance compared with a single model (Figure 5.1).

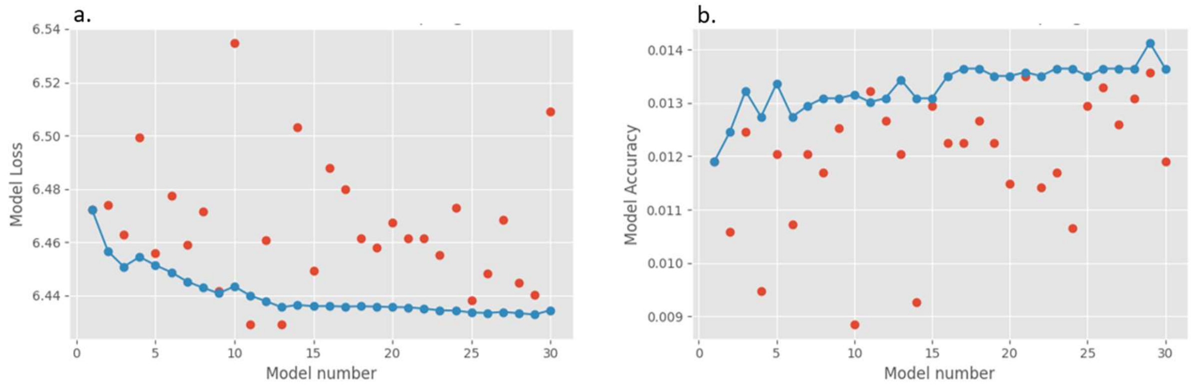
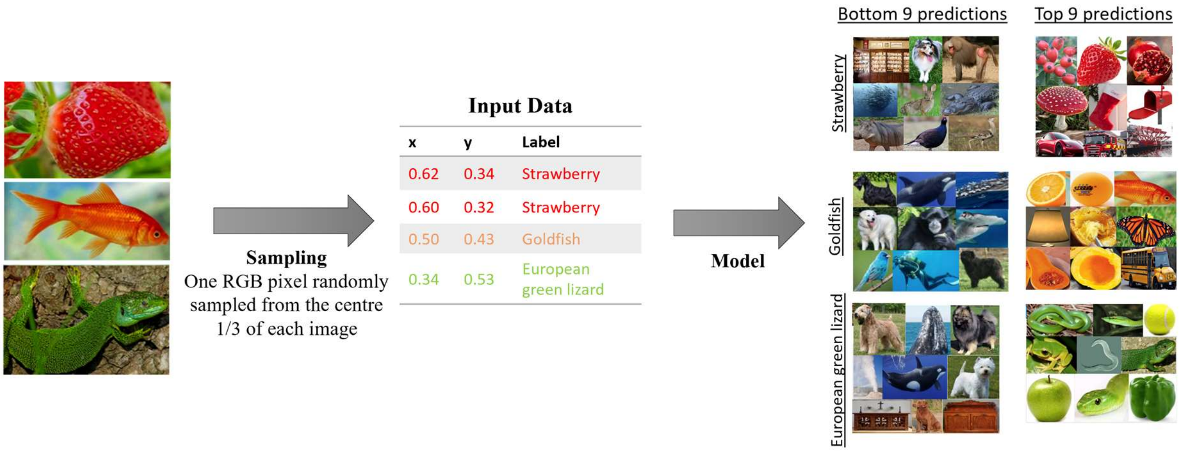


Figure 5.1 *Model performance as additional neural networks are included in the ensemble. (a) Line plot showing single model loss (red dots) and loss of ensembles of increasing size (blue line). (b) Line plot showing single model accuracy (red dots) and loss of ensembles of increasing size (blue line).*

The model was evaluated using colours in an RGB triangle in CIE1931 chromaticity space. The model produced a probability distribution for each colour which gives the estimated probability of each object given the colour (Figure 5.2). Bayes' rule (Equation 5.1) was applied to these probability distributions P(object|colour) to estimate the probability of each colour given the object identity P(colour|object).

Equation 5.1 *Bayes' Rule for P(colour|object).*

$$P(colour|object) = \frac{P(object|colour) \cdot P(colour)}{P(object)}$$



Input Data

x	y	Label
0.62	0.34	Strawberry
0.60	0.32	Strawberry
0.50	0.43	Goldfish
0.34	0.53	European green lizard

Figure 5.2 *Diagram showing example of input and output of a deep neural network trained to identify objects using a single colour.*

5.4 Results

The ensemble of 30 neural network models predicted the correct object class on test data with 1.36% accuracy. Although estimates were far from accurate, they were significantly more accurate than random (0.11% accuracy). At 1.36% accuracy, the model was over 12x more accurate than random, suggesting that valuable information about the objects was learnt.

The relatively low observed accuracy rate was expected, as the task of identifying an object based on its colour alone is very difficult. For example, given a red colour, one might predict that the associated object is a strawberry or a phone box. Although estimating the right object on the first guess is very unlikely, certain possibilities can easily be excluded, for example a whale or a dog. This is how the model performed, by organising objects based on their most probable colours. The most likely object may not be correct, but the most likely objects all possess the same colour properties as the test colour (as illustrated in Figure 5.2).

The probability distribution of colours given an object *P(*colour|object) was plotted on the CIE1931 chromaticity space in Figure 5.3. The distributions vary smoothly, and peak where expected for the respective objects. However, unlike the KDE model, model predictions are not fit to the distribution of the training data, but rather represent the colours which are most useful for recognising that object among other objects. For example, black is a common colour in "Ladybug", but red is the most useful colour for "Ladybug" recognition. As such, the model learnt the colours which are most useful for recognising an object among a group of objects, which can be applied to make predictions about new data.

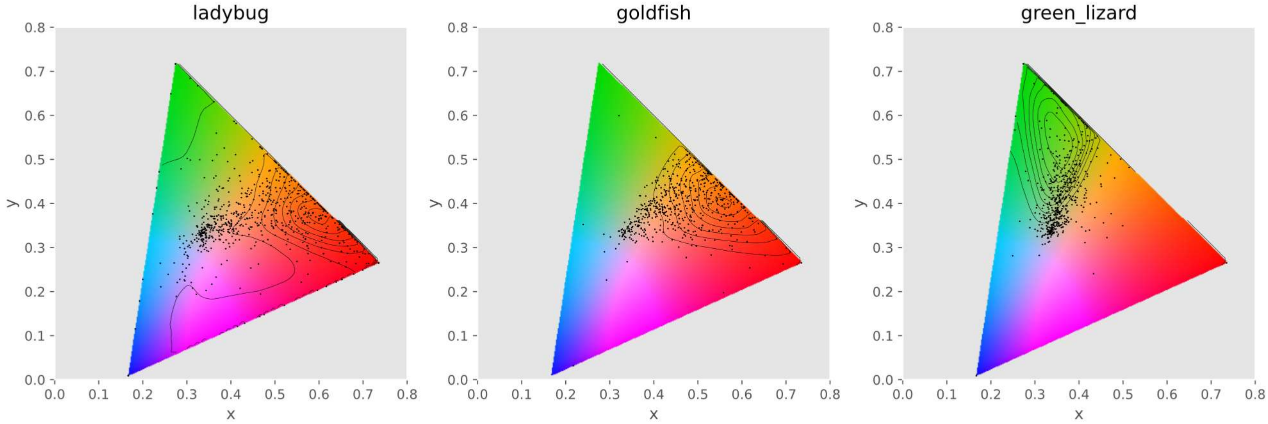


Figure 5.3 *Probability distribution of colours which predict an object class P(colour|object) for object classes: "Ladybug", "Goldfish" and "Green Lizard". Scatterplots show the observed data for that object, and contour plots show model predictions.*

Next, the mutual information between object classes and colours was measured across the colour space using information theory, as in Section 4.5. Mutual information quantifies the amount of information obtained about one random variable, by observing another random variable.

Table 5.1 *Mutual information between object classes and colours for an ensemble neural network.*

	Neural network (1.36% accuracy)
Uncertainty about colour, *H*(colour)	11.60 bits
Uncertainty about colour given object class, *H*(colour|object)	11.18 bits
Average amount of information provided by objects about colour, *I*(colour|object)	11.60-11.18 = 0.42 bits
Uncertainty about object class, *H*(object)	9.79 bits
Uncertainty about object class given colour, *H*(object|colour)	8.85 bits
Average amount of information provided by colour about objects, *I*(object,colour)	9.79-8.85 = 0.94 bits
Average number of colour categories used to discriminate between objects	$2^{0.94}$ = 1.91

5.4.1 What information do objects provide about colour?

How much uncertainty is there about colour?

Given no information about an object, estimating the object colour from all possible colours is a random guess. Uncertainty about colour is high, as all colours in the colour space are equally likely. The entropy of the list of average probabilities of each colour *H*(colour) was 11.60 bits (Table 5.1).

How much uncertainty is there about colour, given the object identity?

To measure uncertainty about colour given the object identity, the entropy of the probability distribution of colours was calculated for each individual object $H(\text{colour}|\text{object})$. The entropy was averaged across all objects, to measure the average uncertainty about colour once object identity is known, which was 11.18 bits (Table 5.1).

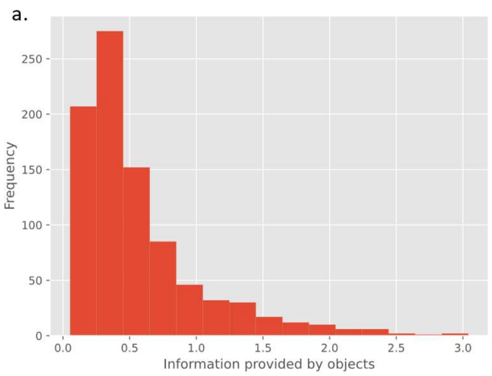
How much information do objects provide about colour?

The difference between uncertainty about colour given no information about the object, and uncertainty once object identity is known, is the amount of information provided by objects about colour (Equation 5.2). Figure 5.4a shows the distribution of information values across different objects. Information was positively skewed, with a median of 0.42bits. Therefore, on average, objects provide 0.42bits of information about colour (Table 5.1).

Some objects provided more information about colour than others (Figure 5.4b, from a minimum of 0.05 bits for "Plunger" up to a maximum of 3.04 bits for "Lemon". This mutual information is a measure of colour diagnosticity, where the more information the object provides about colour, the more colour diagnostic the object is.

Equation 5.2 *Amount of information provided by objects about colour.*

$$I(colour, object) = H(colour) - H(colour|object)$$



b.

Most colour diagnostic objects	Information (bits)	Least colour diagnostic objects	Information (bits)
Lemon	3.04	Plunger	0.05
Rapeseed	2.88	Goblet	0.07
Sulphur butterfly	2.66	Hatchet	0.07
Tiger shark	2.53	Corkscrew	0.08
Hammerhead	2.47	Birdhouse	0.08
Jackfruit	2.42	Hook	0.08
Orange	2.38	Pitcher	0.09
French loaf	2.37	Tobacco shop	0.09
Granny smith	2.33	Spatula	0.09
Electric ray	2.27	Scale	0.10

Figure 5.4 *Information provided by objects about colour for an ensemble neural network. (a) Distribution of information provided by objects about colour, across objects. (b) Table showing the 10 objects which provide the most information about colour (most colour diagnostic) and least information about colour (least colour diagnostic) with the associated information value in bits.*

5.4.2 What information do colours provide about objects?

How much uncertainty is there about objects?

Given a task of identifying one object out of 883 equally possible objects, with no information about that object, there is a 1/883 chance of correctly estimating the object. It would require log2(883)=9.79 yes/no questions to get the correct object. This number is uncertainty about object identity from a uniform distribution of 883 object classes.

The entropy of the list of average probabilities of each object *H(*object) was 9.79, showing that all objects are equally as likely. This is the maximum uncertainty about objects.

How much uncertainty is there about objects, given the colour?

Uncertainty about object identity given the object colour was then measured by calculating the entropy of the probability distribution for each individual colour H(object|colour). If uncertainty changes across the colour space, then this would suggests that some colours are more informative about objects than others. The entropy was averaged across all colours to estimate average uncertainty about object class given object colour, which was 8.85bits (Table 5.1).

How much information does colour provide about objects?

The difference between uncertainty about object identity given no information about the object *H(*object), and uncertainty once colour is known H(object|colour), is the amount of information provided by colour about objects (Equation 5.3). The average amount of information provided by colour about objects was 0.94bits (Table 5.1). This totals to 1.91 colour categories, while it was predicted that colour would provide the same amount of information as six colour categories.

Equation 5.3 *Amount of information provided by colour about objects.*

$$I(object, colour) = H(object) - H(object|colour)$$

Amount of information provided by colour about objects was then plotted on a contour plot over the CIE1931 RGB triangle (Figure 5.5a-b). The information values form peaks and troughs across the colour space. Low saturation colours were least informative about objects, reaching a minimum at the achromatic category focus (x = 0.34, y = 0.35). There is a peak in information for pink (x = 0.35, y = 0.19), red (x = 0.58, y = 0.26), blue (x = 0.21, y = 0.33) and green (x = 0.39, y = 0.53). There is also a smaller peak around orange (x = 0.45, y = 0.49), but it is not as prominent as there is no boundary between green and orange. Finally, there is a drop in information along the Planckian Locus which follows an axis from blue to yellow (Figure 5.5c). Table 5.2 shows the top 10 object class predictions for the five category foci identified by the model in the information geometry. The object classes correspond with the object colours.

Table 5.2 *The top 10 object class predictions for category foci. The column header indicates the chromaticity values of the category foci in CIE1931 chromaticity space. Predictions are listed from most probable object on row 1 in a descending order. The background colour indicates the hue and saturation of the category foci, but the luminance is variable.*

Top predictions	**Pink** **x=0.35, y=0.19**	**Red** **x=0.58, y=0.26**	**Blue** **x=0.21, y=0.33**	**Green** **x=0.39, y=0.53**	**Grey** **x=0.34, y=0.35**
1	Cardoon	Strawberry	Scuba diver	Granny smith apple	Cabbage butterfly
2	Feather boa	Pomegranate	Hammerhead	Green snake	Seashore
3	Velvet	Paddlewheel	Electric ray	Tennis ball	Ice bear
4	Poncho	Christmas stocking	Dugong	Tree frog	Stone wall
5	Jellyfish	Theatre curtain	Sea snake	American chameleon	Sandbar
6	Flatworm	Mailbox	Tiger shark	Bell pepper	Snow leopard
7	Pinwheel	Hip	Street sign	Vine snake	Stove
8	Artichoke	Fire engine	Stingray	Green lizard	Sulphur-crested cockatoo
9	Stole	Jinrikisha	Oscilloscope	Green mamba	Maze
10	Slot	Sports car	Great white shark	Damselfly	Trailer truck

Figure 5.5e-f show the biases in colour perception observed in Chapter 3, where red regions are areas which responses were biased away from (sources), and blue regions were areas which responses were biased towards (sinks). The information geometry for a model trained to recognise objects using colours follows a similar geometry as the biases in colour perception.

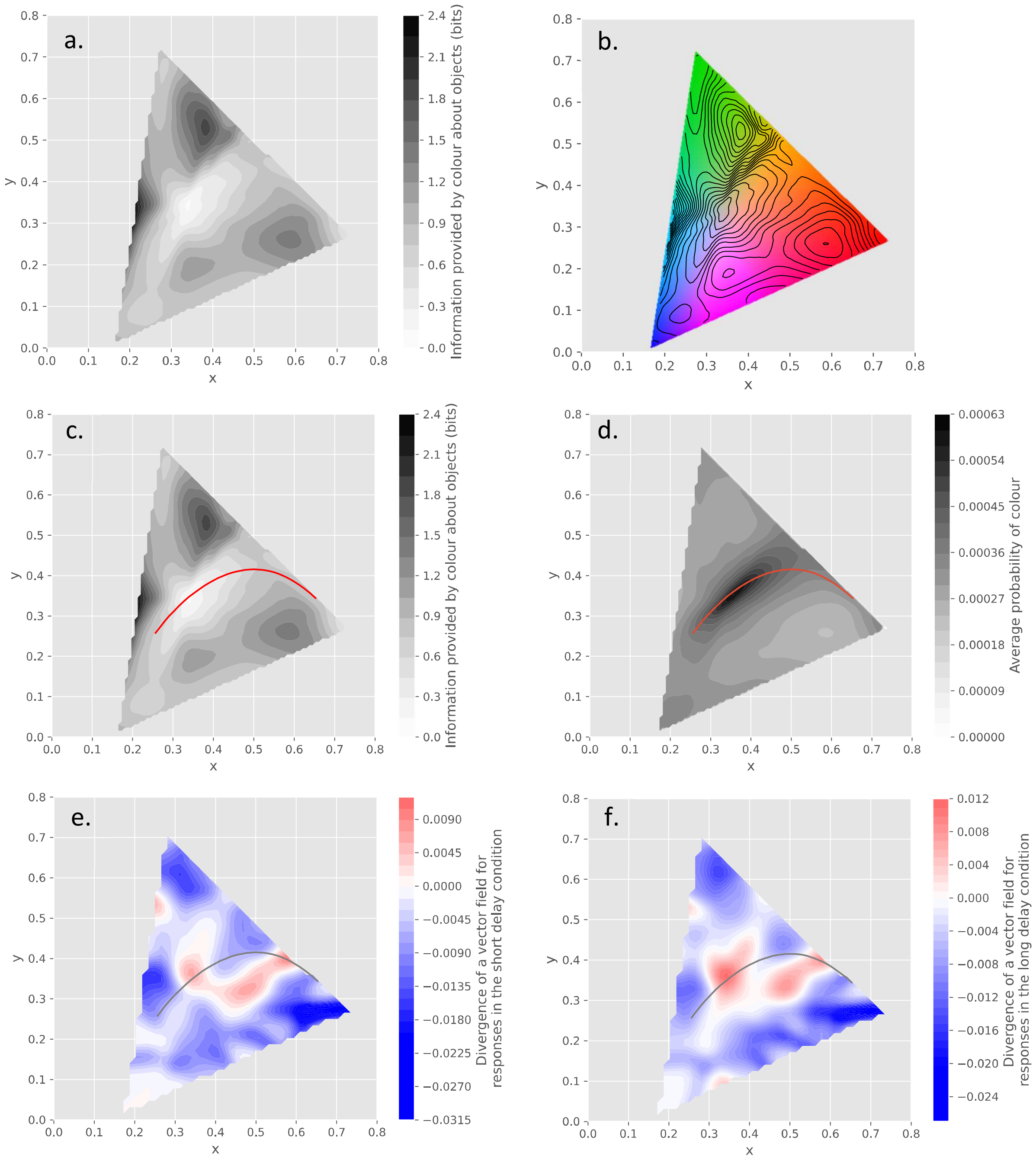


Figure 5.5 *Information provided by colour about objects for an ensemble neural network. (a-b) Amount of information provided by colour about object identity for the current model. Data were binned into 20 levels for both plots. (a) shows the scale of information values, and (b) displays how information interacts with colours in the chromaticity space. (c) Identical contour plot to Figure 5a with the Planckian Locus plotted on top in red. (d) Contour plot showing the average probability of each colour with the model, with the Planckian Locus plotted in red. (e-f) Divergence of a vector field for responses to the behavioural task in Chapter 3 for short delay condition (e) and long delay condition (f). Red areas have positive divergence, therefore they are sources from which responses are repelled. Blue areas have negative divergence and therefore are sinks which attract responses. The grey curved line shows the Planckian Locus.*

5.5 Discussion

It was demonstrated that a deep neural network learns information about how objects are distributed across colour space, and uses this information to predict the probability of objects across colour space. On average, colours did not provide the same amount of information as six colour categories; however the model did organise objects into five distinct colour categories, suggesting that colour categories are useful for recognising objects. Chromatic category foci were more informative about objects than category boundaries, and the achromatic colour category was least informative about objects. The information geometry of object colour space had a similar structure to the geometry of perceptual colour space, such that colour perception was biased towards colours which are most informative about objects.

As predicted, the model successfully organised objects into discernible colour categories which reflect five BCTs: red, green, blue, pink and grey. The chromatic category foci displayed a peak in information, demonstrating that these colours are most informative about objects. The category boundaries between colours are less informative about objects than category foci. There is a trough in information around low saturation colours, reaching a minimum for the achromatic category foci (black/white/grey). This suggests that achromatic colours are less informative about objects than chromatic colours. The saturation effect was also observed in the KDE model (Chapter 4), whereby unsaturated colours are less informative about objects because these colours are more likely. A low saturation colour is not a distinguishing factor to identify an object, as there is a large number of possible objects, so probability estimates for the probability of an object given a low saturation colour are close to uniform.

There is some evidence for a small peak in information around orange (Figure 5.5a-b: x=0.45, y=5), however there is no category boundary between green and orange and therefore the orange category focus is less prominent. The green category focus lies very close to orange and is much larger than the orange category, making it more difficult for the model to discriminate between green and orange objects. It is surprising that the green focus in the information geometry is so close to orange, suggesting that chartreuse (x = 0.39, y = 0.53) is more informative about objects than a standard green (x = 0.3, y = 0.65). This may reflect the distribution of colours across the objects in the ImageNet dataset, as the dataset does not include many plant classes. Plants can have a large effect on the colour statistics of the environment, particularly in wet environments, due to chlorophyll in the leaves reflecting green light (Webster et al., 2007). Therefore some green colours which occur in the environment may be underrepresented in the dataset, resulting in the green category focus which is closer to orange. This hypothesis can be tested by replicating the study with an alternative dataset which includes more flora.

The most remarkable finding in this model is that the information geometry of object colour space varies in a similar structure to the geometry of perceptual colour space, whereby colour perception was biased away from colours which are not informative about objects and towards colours which are most informative about objects. Colour perception was biased away from unsaturated colours, and unsaturated colours were the least informative about objects. In addition, colour perception was biased towards colour categories for green, orange, red, pink and blue, and category foci for green, red, pink and blue (and some evidence of orange) were the most informative about objects. Finally, colours along the Planckian Locus were observed to be less informative, forming a boundary between warm and cool colours. Similarly, colour perception was biased away from the boundary between warm and cool colours, as shown by the red regions in Figure 5.5e-f. These results support the hypothesis that the non-uniform categorical prior which biases colour perception originates from the distribution of colours in the environment, and serves to optimise object recognition, such that colour perception is biased towards colours which are most informative about object identities.

Since the neural network used information about colour to recognise objects, these findings support the notion that humans use colour information for object recognition tasks, and thus the human visual system is adapted to the colour statistics of objects in the environment in order to improve object recognition. For the task of object recognition, an optimal representation of colour should maximise differences between object classes, while minimising differences within classes. These findings suggest that category boundaries between BCTs are useful for discriminating between different objects in the environment, and thus these colour categories are observed in languages universally because they are useful to inform humans about objects.

5.5.1 Effect of colour temperature on informativeness of colours

There is evidence for a category boundary in informativeness between warm and cool colours. The Planckian Locus follows an axis from blue to yellow, and reflects variation in chromaticity as a result of natural daylight. As such, the Planckian Locus echoes the colour statistics of the environment, such that colours lying close to this axis are more common and therefore less informative about objects (Figure 5.5c-d). The Planckian Locus also divides colours into warm and cool, where colours above the axis are cool (blue and green), and colours below the axis are warm (red, orange and pink). Colours lying on the Planckian Locus were less informative, therefore colours on the boundary between warm and cool are less informative than warm and cool colours.

Other researchers have also identified an effect of colour temperature on the informativeness of colours. Gibson et al. (2017) used a communication game to measure how efficiently a Munsell colour chip is communicated using languages in the WCS (Cook et al., 2005). Given a colour term, the higher the accuracy with which the colour chip is identified, the more efficiently the colour chip is communicated. Gibson et al. (2017) observed that all 110 languages in the WCS were optimised for more precise communication of warm colours. There are more BCTs to describe warm colours (red, pink, orange, yellow, brown) than cool (blue and green), hence warm colour chips are communicated more efficiently. Conway et al. (2020) replicated these findings with independently collected colour naming data for three languages: Tsimane (a non-industrial small-scale Amazonian society), Bolivian-Spanish, and English. Across all three languages, red and orange colour chips were always communicated the most efficiently, while blue and green colour chips were always communicated the least efficiently.

To identify the behavioural relevance of warm versus cool colours, Rosenthal et al. (2018) analysed the colour statistics of a database of over 20,000 images of objects. Human observers drew segmentation masks around the objects in the images, and the researchers compared the colour statistics of objects versus background. For both natural and artificial objects, backgrounds tended to be cooler coloured (blue, green and grey), and objects tended to be warmer coloured (red and yellow).

Subsequently, Conway et al. (2020) concluded that communication efficiency is better for warmer colours, because objects are warmer coloured than backgrounds. Therefore, the colour lexicon is optimised to communicate about objects, because objects are more behaviourally relevant than backgrounds. Conway et al. (2020) argued that shape information is primarily used to identify objects, and colour information signals whether the object is behaviourally relevant, and therefore identifies which objects to direct attention towards. For example, the shape of a banana is used to identify the object, and the colour signals the behavioural relevance of the banana (whether it is ripe enough to eat). As a result, colour information supports a more efficient allocation of object recognition resources, such that resources are allocated to the most behaviourally relevant objects, to reduce cognitive load.

If this hypothesis were correct, then it would be expected that warmer coloured objects are more probable than cool coloured objects in the current study, and warm colours are more informative about objects than cooler colours. However, there was no evidence for this hypothesis: warm and cool colours were equally as probable and information was organised into the same number of colour categories for warm and cool colours (Figure 5.5c). Contradictory to arguments by Conway et al. (2020), it was observed that colour categories are useful for identifying objects, with neutral toned colours (along the Planckian Locus) providing the least information about objects, and warm and cool colours provided the most information about objects.

The discrepancy between the colour statistics observed here and colour statistics in Rosenthal et al. (2018) is likely a result of the dataset used to measure colour statistics and the definition of an object. Objects were determined in Rosenthal et al. (2018) by the decisions made when creating the image dataset, where observers manually selected images from image search engines for the purpose of detecting salient objects (Liu et al., 2011). Objects were defined as something that can be grasped. Observers selected images which contained “a clear, unambiguous object of interest” that was either a salient object or a distinctive foreground object, and any “very large” salient objects were excluded (Liu et al., 2011).

The images used are therefore not random natural scenes, but rather determined by the decisions which led to the image selection. Rosenthal et al. (2018) assumed that these decisions reflect the processes of object recognition, and therefore used them as the statistical basis for object judgements. However, object recognition does not occur under perfect viewing conditions, where the object is in the foreground and salient. In addition, the behavioural relevance of an object is dependent on the task. Therefore, it is entirely possible that for a task where participants were asked to select images of salient objects, participants tended to select warm coloured objects which contrast with cool backgrounds. As a result, the study only demonstrates that objects were warmer than backgrounds for images of objects which were specifically selected to only include salient objects.

On the other hand, the ImageNet dataset used in the present study queried 1000 common nouns from WordNet on internet search engines to collect images (Russakovsky et al., 2015). Therefore, in this study an object may be defined as a common noun which can be viewed. Common nouns include objects which cannot be grasped, and therefore would not be classified as an object by Rosenthal et al. (2018), for example ImageNet includes classes for cliff, grocery store, volcano, cinema, church, toy shop etc. Therefore, some of the objects in this study would be classified as background in Rosenthal et al. (2018) or simply excluded from the dataset because of size or saliency. This might explain why warm colours were not more probable than cool colours for objects in ImageNet.

However, warm colours were still communicated more efficiently than cool colours across many languages and multiple studies (Conway et al., 2020; Gibson et al., 2017; Zaslavsky et al., 2019). In this study, two warm colour categories (red, pink) and two cool colour categories (blue, green) were identified, so warm colours did not provide more information about objects than cool colours. There are more BCTs for warm colours than cool colours, but these colour terms do not exist in the chromaticity space used. The orange category was not prominent, but it may show up when using a different dataset. In addition, if the luminance channel is explored, then it is possible that colour categories would be observed for yellow and brown, so there would be more warm colour categories than cool, resulting in warm colours providing more information about objects than cool colours.

5.5.2 Limitations of a two-dimensional colour space

Luminance information was removed because luminance is heavily influenced by illumination, and without any cues about the illuminant (due to a single pixel being sampled from the image), the luminance is not a reliable signal of surface reflectance (Kingdom, 2011). Chromaticity is also influenced by illumination, but to a much lesser extent. The Planckian Locus shows the variation in chromaticity as a result of daylight. For colours along this locus, an object's surface chromaticity is a less reliable signal of surface colour, as chromaticity may be a result of illumination. As such, colours along the Planckian Locus are less informative about object identity because, without any cues to the chromaticity of the illuminant, there is more uncertainty about the object chromaticity. Cues about the illumination, such as relations between surface reflectance in different spatial locations, may resolve some of the uncertainty. However colour constancy is not perfect, so there will always be more uncertainty about object colour along the Planckian Locus.

It was predicted that if the model organises objects into six colour categories, then on average colour would provide log2(6)=2.58 bits of information, as this is entropy of a uniform distribution of six colour categories. In this model, on average colour provided 0.94 bits of information about object identity. However, the prediction was based on the assumption that there would be 6 category foci, which are all maximally informative about objects, and that all 6 colour terms are equally probable. As luminance information was excluded, some colour terms occupy the same area in the chromaticity space - for example black, white and grey - which resulted in a non-uniform distribution of colour categories across the space – low saturation colours were more probable than saturated colours. As a result, the distribution of information values across category foci is not uniform - colours in the achromatic colour category were less informative about objects than colours in chromatic categories. Uniform distributions have maximum entropy, so the entropy is lower than the maximum for a non-uniform probability distribution. Therefore, the average information provided by colour about objects was less than the entropy of a uniform distribution of six colour categories.

Although this model suggests that achromatic colours are not very informative about objects, they are still communicated efficiently since behavioural relevance is also determined by the probability of colours. Conway et al. (2020) observed that across all languages in the WCS there was no correlation between colour chip saturation and communication efficiency of colours, which suggests that low saturation colours are also communicated efficiently.

Baddeley and Attewell (2009) observed that a lightness lexicon with three terms (white, black and grey) was optimal to communicate efficiently about reflectances in the environment. These three colour terms are the three BCTs for achromatic colours. This demonstrates that colour lexicons do contain the optimal colour terms to efficiently communicate about achromatic colours, despite achromatic colours providing little information about objects in the current model. Future research will need to measure the amount of information provided by colour across all three dimensions of colour, in order to assess whether 11 BCTs is optimal to communicate about objects in the environment, without the probability of a colour term confounding the information geometry. This is a complicated task, as it will require modelling lightness constancy whilst excluding shape information.

The amount of information provided by colour about objects was also smaller in this model than in the KDE model in Chapter 4. The KDE model predicts p(colour|object), and the neural network predicts p(object|colour). As a result, probability distributions for colour-concept associations are more variable across objects than across colours, therefore objects provide more information about colour than colours provide about objects.

5.5.3 Colour diagnosticity

The neural network also organised objects by colour diagnosticity, i.e. according to whether colour is a useful property of the object to distinguish it from other objects. The most and least colour diagnostic objects are sensible, where colour diagnosticity seems to reflect how strongly associated an object is to a particular colour (Figure 5.4b). For example, high colour diagnostic objects include many natural objects which are reliably coloured, such as lemon, orange and hammerhead shark. In contrast, low colour diagnostic objects include many artefacts with unreliable colour properties, such as plunger, hook and spatula. This measure of colour diagnosticity seems to more accurately reflect intuitive beliefs about the association between colour and objects than the predictions for the KDE model, and therefore may provide a better measure of colour diagnosticity. Although a combination of the information provided by colour about objects and the information provided by objects about colour may be valuable, and could help to resolve some of the contradictions around defining and measuring the colour diagnosticity of objects (Tanaka & Presnell, 1999). The global assignment hypothesis suggests that when making associations between colours and objects, decisions will be optimised based on the entire set (Schloss et al., 2018). As such, colours which are not associated with an object are also useful for object identification.

However, it is important to note that colour diagnosticity, measured by how informative colour is about objects compared with objects, as used in this model, is dependent on the other types of objects in the model. This model contains a very large range of objects, including vehicles, artefacts, buildings, plants and animals. Plants and animals tend to be more colour diagnostic that artefacts, vehicles and buildings, as man-made objects can be artificially coloured, resulting in more variation in colours across instances of the same object class. On the other hand, natural objects have more reliable colour properties due to biological constraints on colouration. In addition, colour may be more useful for discriminating between objects which have similar shapes, for example different species of an animal.

5.5.4 The dataset

As such, the objects within the image dataset used to sample the object colours has a large impact on the amount of information that colours provide about objects. The real world contains an infinite number of objects, but to test the theory, the current methods are limited to a dataset which has a finite number of objects. Not only is the number of objects important, but also the type of objects. Object categories are organised in hierarchies, where some categories are superordinate, and others are subordinate, e.g. bear, versus types of bear – panda, polar, koala etc. (Rosch, 1978). The ImageNet dataset follows the WordNet hierarchy, but all of the categories are not of the same level of abstraction. For example, there may be many species of one animal, but a generic term for another animal. This is of importance for the present study, because the task given to the neural network (to identify an object using object colour), is intended to represent a real world task where humans use colour to improve object recognition. As such, the object categories included in the dataset should reflect the types of objects which humans are likely to discriminate between.

Since the task is so dependent on the types of objects in the dataset, it would be valuable to test this model using a different image dataset to see if similar patterns of information are observed across colour space. It is possible that using different types of object classes (e.g. different hierarchies) changes the way information about objects varies across colour space. In addition, if a certain class of objects is underrepresented or overrepresented in the dataset compared with its occurrence in the environment, then this may influence how informative colour is across all objects. For this reason, it is important to test whether a model trained using a different dataset also produces an information geometry which reflects the geometry of perceptual colour space. This would suggests that the information geometry observed in this chapter represents the colour statistics of the environment, not just properties of the ImageNet dataset.

5.5.5 Summary

It was observed that a deep neural network learned the information that colours provide about objects and used this information to group objects into five colour categories: green, red, pink, blue and grey. The information geometry of the object colour space reflects the geometry of perceptual colour space, such that colour perception was biased towards the colours which are most informative about objects, and away from colours which were least informative about objects. These results suggest that the distribution of colours across objects may act as a prior, where uncertainty about sensory signals is reduced by directing perceptual processes towards conclusions which are most informative about objects. This model is dependent on the object classes included in the training data, with an underlying assumption that humans use colours to discriminate between these objects. While ImageNet includes a large number and variety of object classes, it would be valuable to test whether the findings can be replicated with a different dataset. If the findings are replicated, then this would provide further evidence that the information geometry across object colour space reflects the distribution of colours in the environment, and forms the basis of a prior in high-level colour perception.

6 The amount of information provided by colours about object identity using a deep neural network: a replication study.

A deep neural network was trained to identify objects using colours sampled from objects in the Open Images V6 dataset, to test whether the distortions to the information geometry induced by object recognition for a model trained using ImageNet could be replicated with an alternative dataset. The Open Images model organised information provided by colour about objects into seven distinct colour categories: green, orange/yellow, red, pink, blue, beige and grey. The location of the category foci in the Open Images model closely corresponded with the ImageNet model, and therefore the results were successfully replicated. In addition, the information geometry of object colour space followed a similar structure to the geometry of high-level colour perception, such that colours which were most informative about objects were those towards which colour perception was biased, and the colours which were least informative about objects were those which perception is biased away from. Since similar colour categorical structures were observed in the colour statistics of objects from two independent image datasets and in the geometry of high-level colour perception, this provides strong evidence that the distribution of colours across objects in the environment form the basis for a non-uniform prior which directs perceptual processes towards the colours which are most informative about objects.

6.1 Introduction

In Chapter 5, a deep neural network was trained to recognise objects in ImageNet using the object colour. The model predictions were used to measure the amount of information provided by colours about objects. The neural network organised objects into colour categories, such that the information geometry formed clusters around the BCTs: green, red, pink, blue and grey. In addition, the information geometry of object colour space reflected the geometry of perceptual colour space, such that colour perception was biased towards the colours which are most informative about objects, and away from colours which are least informative.

The goal of this study was to replicate the findings in the previous chapter using an alternative dataset, to test whether the observed information geometry is possibly a feature of the environment or just a feature of the ImageNet dataset. This is important, because the neural network learns to discriminate between all possible objects using colour, and therefore the object classes included in the dataset have a large impact on how informative colour is for object recognition.

It was predicted that a similar information geometry would be observed with a different dataset if the observed patterns in the data represent the statistics of colour in the environment. Therefore, it was predicted that a model trained using a different dataset would also organise information provided by colour about objects into at least five distinct colour categories – green, red, pink, blue and grey – and that chromatic colours would be most informative about objects, with achromatic colours least informative.

There were three factors that were important when searching for an alternative image dataset, to allow a fair comparison with ImageNet: 1) there must be a wide variety of object classes, 2) there must be bounding box annotations, 3) there must be enough data in each object class for the machine to learn the colour statistics of the object. Given these criteria, there was one open source image dataset identified that was acceptable for the object recognition task: Google Open Images V6.

6.2 Methods

The images used were sampled from Google Open Images V6 (Kuznetsova et al., 2020). This open source dataset consists of 9 million images in total, which were collected by sampling all of the images on Flickr, and removing images that contain inappropriate content (e.g. pornography, violence, memes), images that were duplicates, and images that appeared elsewhere on the internet.

Image-level labels were applied using computer-generated, human-verified classes from a list of 19,794 classes ranging across coarse-grained object classes, fine-grained object classes, scenes, events, materials and attributes. From the 19,794 image-level classes, 600 object classes were selected by Kuznetsova et al. (2020) as the most important and definitively boxable, to generate bounding box annotations. These 600 classes included animals, clothing, food, vehicles, buildings, people, sports equipment, kitchenware and furniture. In total there are 1.9 million images containing bounding box data for these 600 classes.

90% of the bounding boxes were human generated, and 10% of bounding boxes used human verification for automatically produced bounding boxes. For human generated bounding boxes, participants were instructed to draw the smallest possible box that contains all visible parts of a given object for all instances of the object in the image. Therefore a single image contains multiple bounding boxes for different objects and repeated instances of objects, so for each image there are on average 8.42 bounding boxes. Before annotating images for each object class, participants were shown good and bad examples in order to improve consistency of annotations.

The 600 object classes for bounding box data, were organised into a hierarchy, for example furniture, table and coffee table are three levels of abstraction. Superordinate categories (e.g. furniture) are much more inclusive than subordinate categories (e.g. coffee table; Rosch et al., 1978). To ensure that there is no overlap in meaning between categories, any object class which had lower levels of abstraction were removed from the dataset (see Figure 5 in Kuznetsova et al. (2020) for a flare dendrogram showing the hierarchy of the 600 boxable classes). Controlling the levels of categorisation prevents any confusion in the object recognition task, for example discriminating between whether an object is “Furniture”, “Table” or “Coffee table”, when in fact all three could be true. By removing the labels “Furniture” and “Table”, the correct object class is less ambiguous. 73 superordinate categories were removed.

Forty classes which did not contain a minimum of 100 bounding box annotations were removed, as these classes do not have enough data for the machine to learn information about the object’s colour properties. Furthermore, a maximum of 10,000 bounding boxes were randomly sampled for inclusion from any classes with more than 10,000 bounding boxes, in order to reduce processing time. Images which were marked as depictions of the object, for example a cartoon or a drawing, were also removed. In total, 1394941 bounding boxes were used, organised into 487 classes, resulting in a mean value of 2864 data points per class.

The image data were processed using the same method as used for the ImageNet data (Section 4.2). The images were cropped to the bounding box to remove the background.

The cropped image was then split into a 3x3 grid, and a random pixel was selected from the central section of each image. The RGB value of this pixel was used to represent a colour within the object. As such, there was a single RGB value for each image in the database, along with its object classification. The resulting RGB values were converted into CIE1931 xyY chromaticity coordinates using the CIE RGB to XYZ transformation matrix and XYZ to xyY transformation (Lindbloom, 2017). The Y value, representing luminance, was removed, and only xy chromaticity coordinates were included, to represent hue and saturation dimensions of colour space. The training data had 2 features (x and y chromaticity coordinates), 1394941 samples and 487 classes of objects.

The same methods were used as outlined in Section 5.3 to train an ensemble of 30 deep neural networks to recognise objects using colours. Figure 6.1 demonstrates the progressive improvement in accuracy and loss of the model as further neural networks were included in the ensemble.

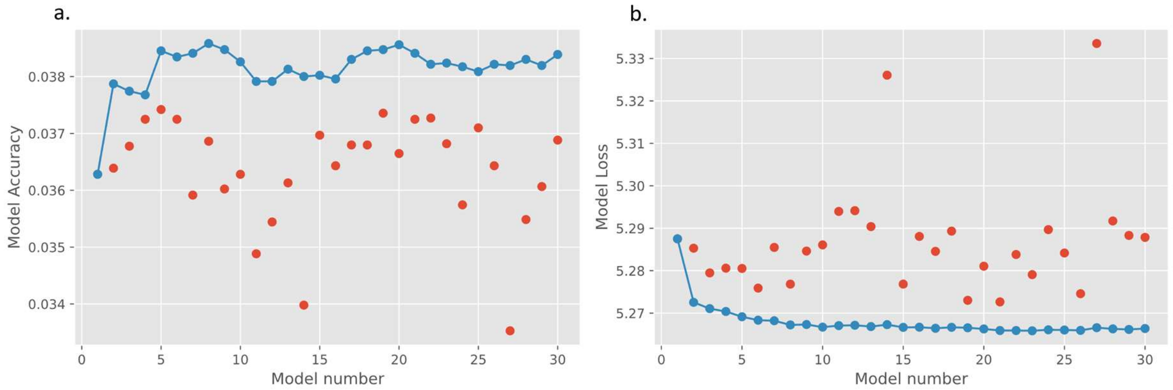


Figure 6.1 *Model performance as additional neural networks are included in the ensemble for the Open Images model. (a) Line plot showing single model loss (red dots) and loss of ensembles of increasing size (blue line). (b) Line plot showing single model accuracy (red dots) and loss of ensembles of increasing size (blue line).*

6.3 Results

The ensemble of 30 neural network models trained to recognise objects in Open Images V6 (Open Images model) predicted the correct object class with 3.84% accuracy (Table 6.1). Although estimates were far from accurate, they were significantly more accurate than random (0.21% accuracy). At 3.84% accuracy, the model was over 18x more accurate than random, suggesting that valuable information about the objects was learnt.

In addition, the Open Images model achieved a higher accuracy rate than the ensemble neural network trained using ImageNet data (ImageNet model).

The probability distribution of colours given an object *P(*colour|object) is plotted on the CIE1931 chromaticity space in Figure 2a. The distributions vary smoothly, and peak exactly where expected for the respective objects. As such, the model learnt the colours which are most useful for recognising an object among a group of objects, which can be applied to make predictions about new data. The probability distributions are similar for comparable object classes in the ImageNet model (Figure 6.2b), but there are also slight differences, for example a larger probability of blue colours for the goldfish class in the Open Images model.

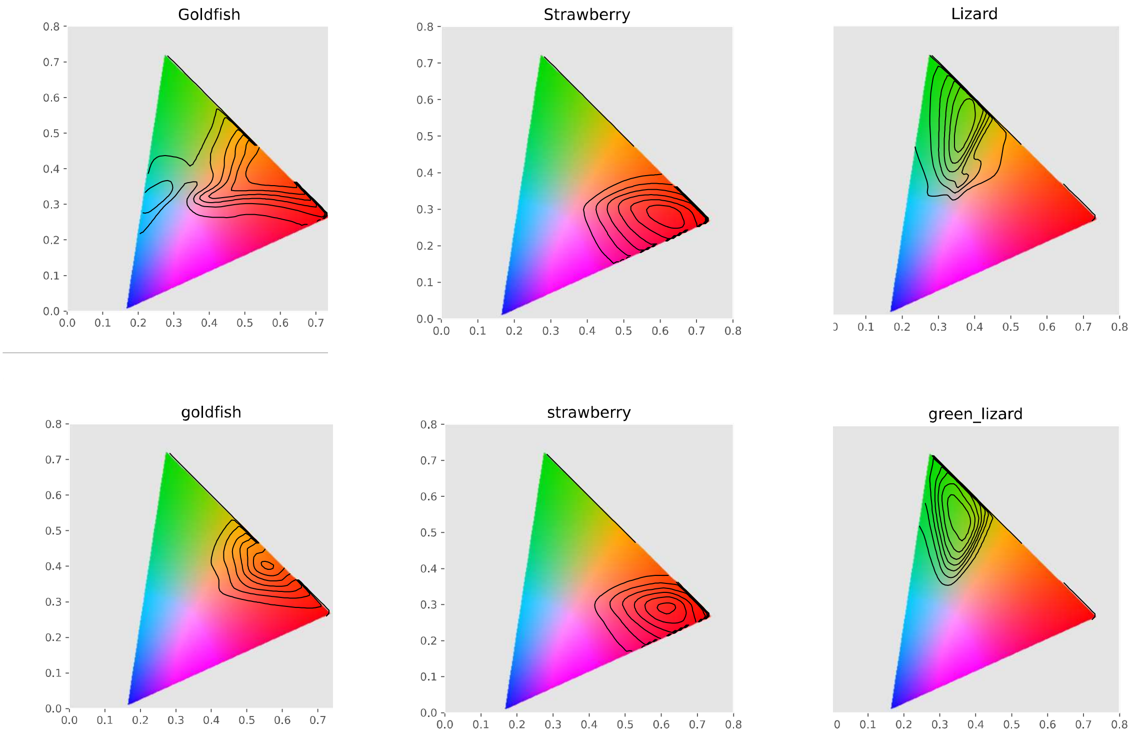


Figure 6.2 *Probability distribution of colours which predict an individual object class P(colour|object) for the Open Images Model (a) and comparable classes in the ImageNet model (b).*

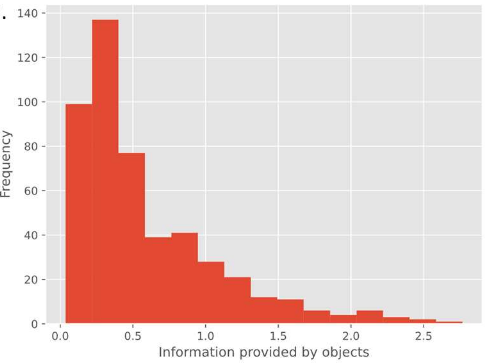
Next, the mutual information between object classes and colours was measured across the colour space using information theory, using the same methods as outlined in Section 5.4.1 and 5.4.2.

Table 6.1 *Mutual information between object classes and colours for an ensemble model of 30 neural networks trained on ImageNet and Open Images.*

	ImageNet Model (1.36% accuracy)	Open Images Model (3.84% accuracy)
Uncertainty about colour, H(colour)	11.60 bits	11.60 bits
Uncertainty about colour given object class, H(colour|object)	11.18 bits	11.18 bits
Average amount of information provided by objects about colour, I(colour|object)	11.60-11.18 = 0.42 bits	11.60-11.18 = 0.42 bits
Uncertainty about object class, H(object)	9.79 bits	8.93 bits
Uncertainty about object class given colour, H(object|colour)	8.85 bits	7.26 bits
Average amount of information provided by colour about objects, I(object,colour)	9.79-8.85 = 0.94 bits	8.93-7.26 = 1.67 bits
Average number of colour categories used to discriminate between objects	$2^{0.94}$ = 1.91	$2^{1.67}$ = 3.18

The amount of information provided by objects about colour was plotted on a histogram (Figure 6.3a). The distribution is positively skewed with a median of 0.42 bits, indicating that objects provided 0.42 bits of information about colour on average, which is identical to the ImageNet model (Table 6.1). Some objects were more informative about colour than others, from a minimum of 0.04 bits for "Chair" up to a maximum of 2.77 bits for "Rays and skates" (Figure 6.3b). This information metric is a measure of colour diagnosticity, where the more information the object provides about colour, the more reliable the object's colour properties are, and thus the object is more colour diagnostic.

a.



b.

Most colour diagnostic objects	Information (bits)	Least colour diagnostic objects	Information (bits)
Rays and skates	2.77	Chair	0.04
Shark	2.44	Mug	0.04
Swimming pool	2.43	Poster	0.05
Tennis ball	2.37	Cart	0.05
Canary	2.36	Bookcase	0.06
Guacamole	2.31	Book	0.06
Lemon	2.18	Stairs	0.07
Submarine	2.17	Kettle	0.08
Broccoli	2.12	Scissors	0.08
Dolphin	2.05	Shelf	0.08

Figure 6.3 *Information provided by objects about colour for the Open Images model. (a) Distribution of information provided by objects about colour, across objects. (b) Table showing the 10 objects which provide the most information about colour (most colour diagnostic) and least information about colour (least colour diagnostic) with the associated information value in bits.*

Overall uncertainty about objects is lower in the Open Images model because it contained less object classes than the ImageNet model. On average, colour provided 1.67 bits of information about objects, so colours were more informative about objects in the Open Images model than the ImageNet model (Table 6.1). However, the Open Images model still did not learn the same amount of information as would be expected to discriminate between 6 colour categories (2.58 bits). Amount of information provided by colour about objects was then plotted on a contour plot over the CIE1931 RGB triangle (Figure 6.4a-b). The information values form local peaks and troughs across the colour space. There is a peak in information for green (x=0.38, y=0.56), orange (x=0.47, y=0.52), red (x=0.63, y=0.23), pink (x=0.30, y=0.20), and blue (x=0.22, 0.38). Low saturation colours were least informative about objects, reaching a minimum close to the achromatic category focus (x=0.35, y=0.36). There is also a small peak around beige (x=0.39, y=0.34). Table 6.2 outlines the top 10 predicted object classes for each category focus.

To assess the models, category foci predicted from the computational models were compared with the distribution of empirical data from the behavioural experiment. The local minima and local maxima of the information geometry were identified as predicted category foci, and plotted over the geometry of perceptual colour space (Figure 6.4e). With the exception of the beige category focus in the Open Images model, the location of local minima and maxima in the information geometry of both object recognition models closely corresponded with the divergence values. Colours which were most informative about objects were those towards which colour perception was biased, and the colours which were least informative about objects were those which perception is biased away from. Furthermore, the peaks and troughs in the information geometry are located around the focal points of colour categories, corresponding to green, orange, red, pink, blue and grey (Figure 6.4d). Therefore, the category foci in the computational models based on colour statistics of objects closely predicted the geometry of high-level colour perception.

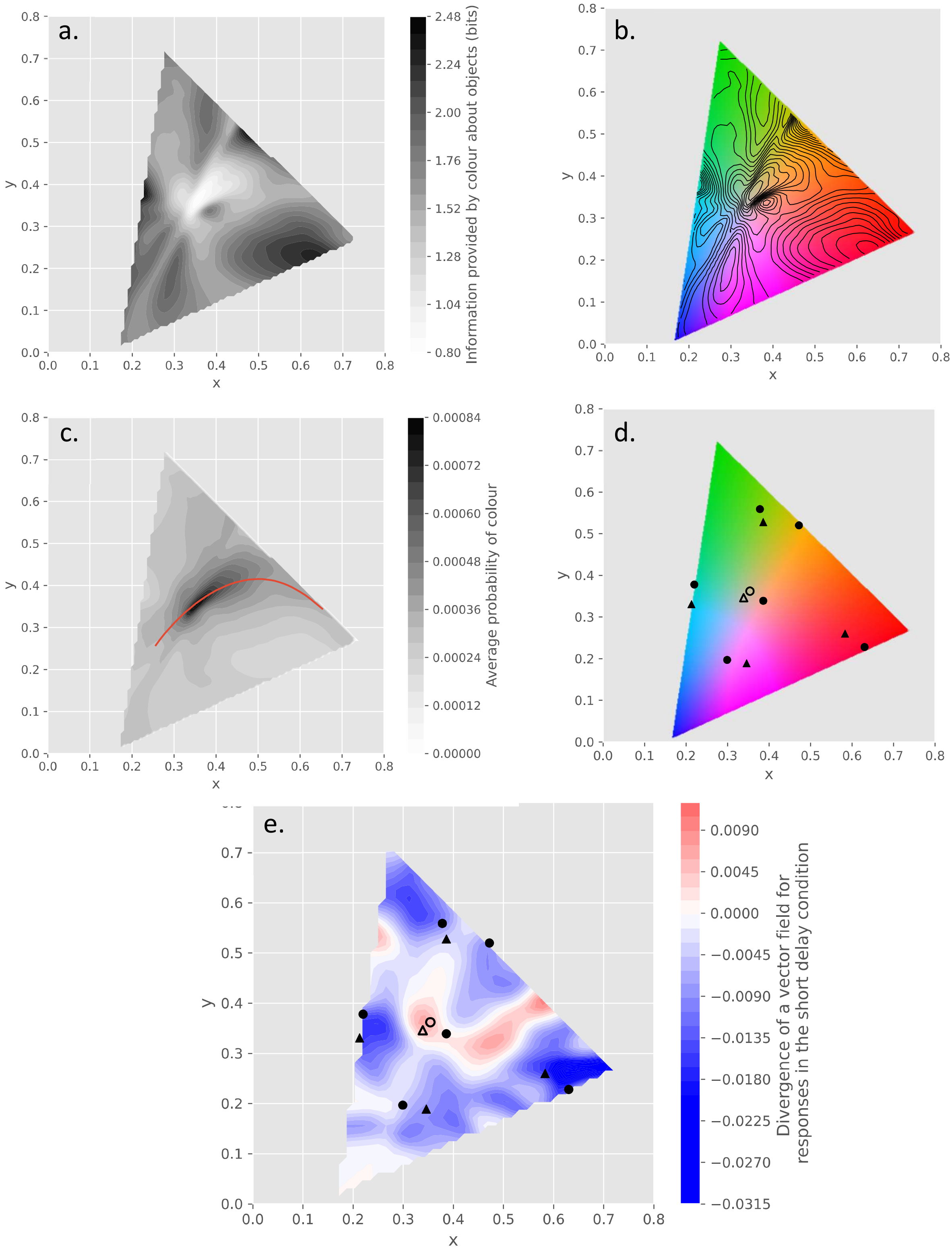


Figure 6.4 *Information provided by colour about objects for the Open Images model. (a-b) Amount of information provided by colour about object identity for the Open Images model. The same data were binned into 20 levels for both plots. (a) shows the scale of information values. (b) displays how information interacts with colours in the chromaticity space. (c) a contour plot showing the average probability of each colour in the chromaticity space, with the red line showing the Planckian Locus.*

(d) The local maxima (filled shapes) and local minima (unfilled shapes) in information geometry for the Open Images model (circles) and the ImageNet model (triangles). This plot demonstrates the colours corresponding with category foci identified in the computational models. (e) A contour plot showing the divergence of a vector field for responses to the colour estimation task in Chapter 3 for the short delay condition. Red areas have positive divergence, therefore they are sources from which responses are repelled. Blue areas have negative divergence, and therefore are sinks which attract responses. The scatterplot overlaying the contour plot shows the local maxima (filled shapes) and local minima (unfilled shapes) in information geometry for the Open Images model (circles) and the ImageNet model (triangles).

Table 6.2 *The top 10 object class predictions for category foci in the Open Images model. The column header indicates the chromaticity values of the category foci in CIE1931 chromaticity space. Predictions are listed from most probable object on row 1 in a descending order. The background colour indicates the hue and saturation of the category foci, but the luminance is variable.*

Top 10 predictions	**Orange x=0.47, y=0.52**	**Pink x=0.30, y=0.20**	**Green x=0.38, y=0.56**	**Red x=0.63, y=0.23**	**Beige x=0.39, y=0.34**	**Blue x=0.22, y=0.38**	**Grey x=0.35, y=0.36**
1	Common sunflower	Lavender (Plant)	Houseplant	Rose	Human ear	Traffic sign	Bee
2	Orange	Drum	Broccoli	Strawberry	Human nose	Balloon	Cattle
3	Lemon	Guitar	Palm tree	Flag	Human face	Traffic light	Picture frame
4	Balloon	Microphone	Salad	Swimwear	Glasses	Billiard table	Book
5	Traffic sign	Balloon	Flowerpot	Balloon	Human hand	Swimming pool	Shelf
6	Vehicle registration plate	Dress	Grape	Dress	Human mouth	Flag	Desk
7	Maple	Human body	Bell pepper	Drum	Human eye	Billboard	Palm tree
8	Beer	Tie	Bottle	Sports uniform	Human head	Waste container	Duck
9	Banana	Girl	Parrot	Guitar	Human hair	Swimwear	House
10	Juice	Goggles	Cucumber	Jacket	Sunglasses	Shorts	Cat

6.4 Discussion

The results from the previous chapter were successfully replicated with an entirely different dataset. A deep neural network trained to recognise objects in Open Images V6 categorised objects into seven colour categories, corresponding to the colour terms orange, pink, green, red, beige, blue and grey. The information geometry of object colour space for the Open Images model had a similar structure to that for the ImageNet model, with peaks and troughs in information occurring in similar regions. In addition, the information geometry of both the Open Images model and the ImageNet model follows a similar structure to the geometry of high-level colour perception, with a bias towards saturated colours, and category boundaries between the foci of BCTs.

6.4.1 Comparison between Open Images model and ImageNet model

Objects provided the same amount of information about colour in both models, despite being trained using different datasets. ImageNet had 883 classes while Open Images only had 487, so the ImageNet classes included a larger range of objects and more fine-grained classes. The hierarchy of object classes was also controlled in Open Images but not in ImageNet. Despite these differences, both models had a positively skewed distribution of information values across objects with an identical median. This suggests that the set of object classes have comparable variation in colour diagnosticity, with a similar proportion of colour diagnostic objects (e.g. fruits, vegetables, animals) and non-colour diagnostic objects (e.g. vehicles, artefacts, buildings).

Colours provided more information about objects in the Open Images model compared to the ImageNet model, despite objects providing exactly the same amount of information about colour. This is a result of six informative chromatic colour categories being identified in the Open Images model compared with four in the ImageNet model. As the information value is higher for chromatic category foci, the existence of more chromatic foci increased the average amount of information provided by colour. This is also reflected in the higher accuracy rate of the Open Images model - the model learnt more information about colours, and therefore more accurately recognised objects using colour.

Despite identifying a total of seven colour categories in the information geometry, the average amount of information provided by colour about objects (1.67bits) suggests that the model only learnt to discriminate between 3.18 categories ($2^{1.67}$=3.18). As explained in the previous chapter, this is due to the colours being collapsed across the luminance channel, resulting in a non-uniform distribution of colour terms across the chromaticity space. The same pattern was observed in the Open Images dataset as the ImageNet dataset, where unsaturated colours were much more common than saturated colours (Figure 6.4c). As such, unsaturated colours are not informative about objects, as the model is unable to discriminate between black, white and grey objects. Since the achromatic category focus forms a local trough rather than a peak in information, the average amount of information provided by colour about objects is less than expected for seven maximally informative colour categories. If luminance information were included in the model to enable identification of individual achromatic foci, then peaks in information might be observed for achromatic colours rather than a trough, and therefore the average amount of information might be equal to the entropy of a uniform distribution of 11 colour categories.

There were two notable differences between the information geometry in the Open Images and ImageNet model. Firstly, a category boundary was not observed between green and orange in the ImageNet model, but was clearly identified in the Open Images model. This suggests that predictions from the Open Images model were more uniform for colours on the boundary between green and orange than the foci of green and orange. One possible reason for this difference is due to the number of category boundaries occurring in this narrow region of the colour space. Due to the lack of luminance information, this region covers the category boundaries between green, orange, yellow and brown.

Looking at the top 10 predictions for objects across the green-orange boundary, the Open Images model seems to divide objects into green and orange/yellow with a boundary at x=0.43,y=0.51, where the model is uncertain about whether the object is green, orange or yellow (Figure 6.5). As such, the model groups orange and yellow objects into the same colour category and identifies a boundary between cool green coloured objects (such as houseplant, broccoli and salad) and warm yellow/orange coloured objects (such as sunflower, orange and lemon). However, in the ImageNet model, the top 10 predictions gradually change from green objects, to yellow, and then to orange (Figure 6.5d). As the green, orange and yellow categories occur in such a narrow region of space, uncertainty about object identity does not reduce between the categories, therefore a clear boundary was not observed in the ImageNet model. Given that the neural network architecture was identical between the computational models, different patterns in the information geometry around green, yellow and orange are a result of differences in the distribution of colours across objects in the datasets.

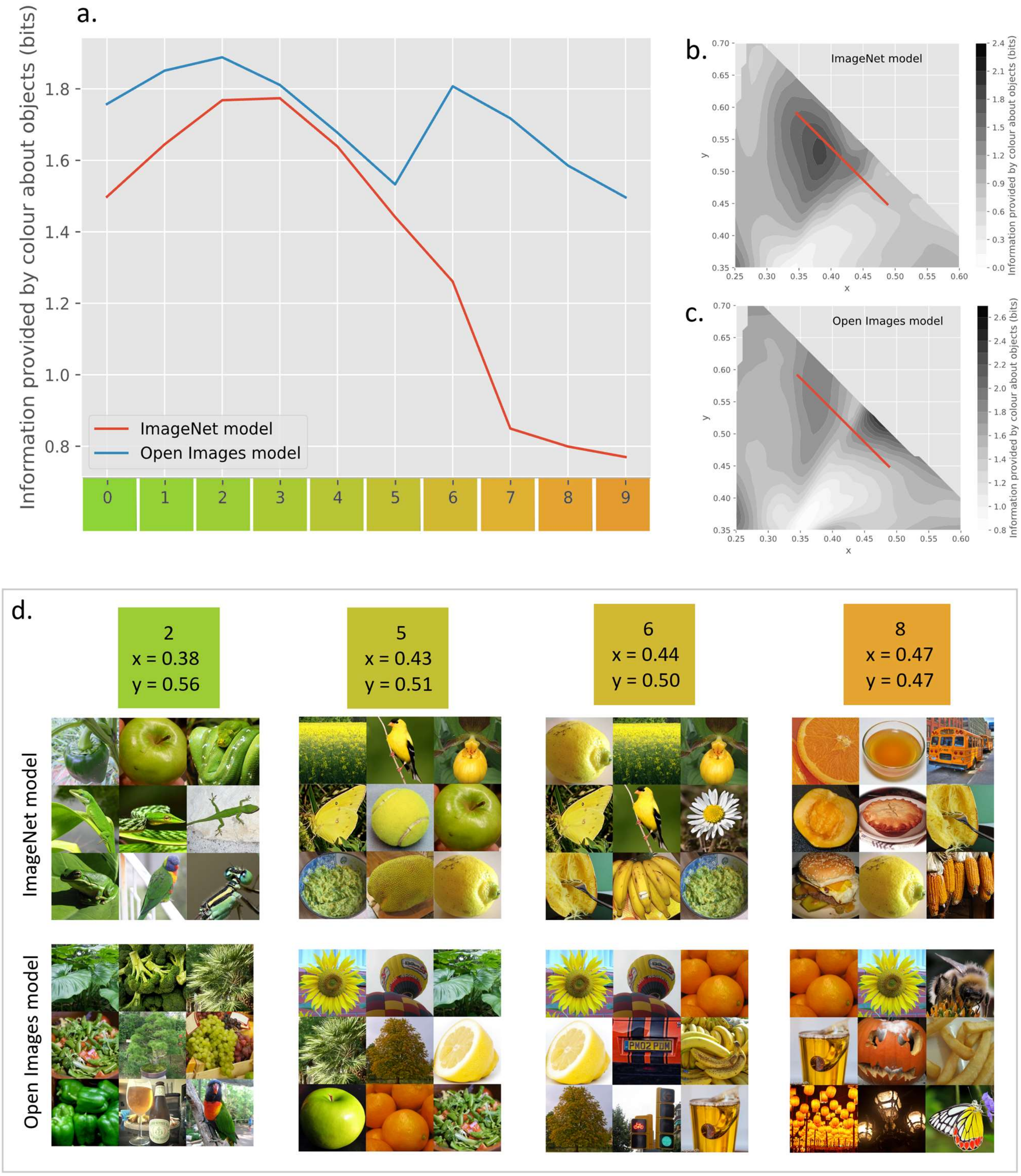


Figure 6.5. *Model predictions across the green-orange category boundary for the ImageNet model and Open Images model. (a) A line plot showing the information value for the ImageNet model and Open Images model for 10 steps in CIE1931 chromaticity space from green to orange. (b-c) The red line shows the region from which the green-orange colours were sampled across the 2D information geometry for each model. (d) Images of the top 9 predicted objects for 4 colours ranging from green to orange.*

The second difference between the two computational models is a small category focus which was observed around beige in the Open Images model but not in the ImageNet model, suggesting that beige colours are more informative than their neighbouring colours for objects in the Open Images dataset. This is due to multiple categories in Open Images corresponding to human body parts, such as human ear, human nose, human face etc. (Table 6.2). Human body parts differ in their shape but, on a single person, the surface chromaticity has little variation (comparitive to the size of the colour space). In addition, variation in luminance enables the inclusion of multiple skin tones into a single category. Therefore the Open Images model has identified this small range of beige colours which vary in luminance as very useful for the recognition of human body parts.

These differences between the models demonstrate that the images sampled and the object classes included in the object recognition task do have an effect on the information geometry of object colour space. The model successfully identifies patterns in the training data, and uses this information when making predictions about objects. This suggests that the colour categories in language reflect systematic patterns in colouration across behaviourally relevant objects in the environment. Therefore, cultural or ecological changes in the behavioural relevance of objects and their associated colours may result in relational changes to colour categorisation. For example, according to Berlin and Kay (1991), languages with five BCTs will have colour terms for white, black, red, green and yellow. Cultural change in the behavioural relevance of blue objects may lead to an increased need to communicate about the colour blue. As a result, usage of blue colour terms may increase among language speakers, until the language adopts blue as an additional basic colour term. As such, the distribution of colours across objects in the environment can explain both universal constraints in colour naming, and variation across languages.

The identification of two additional warm colour categories in the Open Images model (yellow/orange and beige) means that warm colours do provide more information about objects than cool colours. These findings support previous research which identified that warm colours are communicated more efficiently than cool colours across many languages (Conway et al., 2020; Gibson et al., 2017; Zaslavsky et al., 2019). This suggests that there is a universal trend towards more colour terms for warm colours than cool colours as warm colours are more useful for object recognition.

6.4.2 Comparison between the computational models and perceptual model

Finally, both the Open Images model and the ImageNet model follow a similar geometry to the geometry of high-level colour perception (Figure 6.4e). Colour perception was biased away from the achromatic colours and category boundaries which are not very informative about objects, and colour perception was biased towards chromatic colour categories that are very informative about objects. Considering that neither the colour estimation task, nor the object recognition task, required explicit use of colour categories, it is remarkable that a similar categorical structure was identified across the colour space for independent tasks and datasets. This provides strong evidence that the non-uniform distribution of colours across objects in the environment reflect a cognitive prior, which guides colour perception and memory towards colours which are most informative, to improve visual processing and communication about colours. Given that the category foci observed in the colour statistics of objects correspond with BCTs, this further suggests that ecological constraints on colour perception can explain universal patterns in colour categorisation.

There are small variations in the location of foci in the computational models and the perceptual model. For example, the location of the green category focus is closer to yellow/orange in the computational models. This reflects the distribution of green colours across objects in the ImageNet and Open Images datasets, both of which have very few object classes for plants. In the Open Images dataset there are around 100 classes for animals, but the only classes that are related to plants are fruits and vegetables in the context of food. Plants are a large component of the visual environment, and the chlorophyll in plant leaves reflect green light: therefore the full range of green colours experienced in the world are not represented very well in the Open Images model. A model trained using data with a more representative sample of plant classes may predict the green foci in perceptual colour space more accurately. Currently, this method is limited by the datasets available for training the model. However, as more image datasets become available which are representative of the real world, it should be possible to improve the accuracy of the model's predictions of perceptual foci.

In conclusion, it was observed that the distortions to the geometry induced by object recognition was similar for a model trained using ImageNet and Open Images V6. In both models, peaks in information were observed for the chromatic colour categories green, red, pink and blue, and troughs were observed for grey. In addition to the peaks observed in the ImageNet model, the Open Images model identified categorical peaks in information for orange/yellow and beige. Finally, the information geometry of the object colour space in the computational models reflects the geometry of perceptual colour space, such that colour perception was biased towards the colours which are most informative about objects, and away from colours which were least informative about objects. This provides strong evidence that the distribution of colours across objects in the environment act as a non-uniform prior within colour perception, where uncertainty about sensory signals is reduced by directing perceptual processes towards conclusions which are most informative about objects.

7 General Discussion

7.1 Motivation for the project

Universal constraints on colour naming exist, such that 11 BCTs are observed across languages universally (Berlin & Kay, 1991; Cook et al., 2005). However, the origins of these constraints are not well understood (Regier & Kay, 2009; Siuda-Krzywicka et al., 2019; Witzel, 2019). There is evidence for constraints on colour lexicons whereby all languages have the purpose of enabling efficient communication among their speakers (Chaabouni et al., 2021; Regier et al., 2015; Zaslavsky et al., 2018). However, categorical colour perception is observed independently of language, in pre-linguistic infants, animals and patients with brain damage (Siuda-Krzywicka et al., 2019), which suggests that universal patterns in colour categorisation cannot be entirely attributed to constraints arising from language.

The colours that humans perceive and communicate about are strongly associated with objects (Palmer & Schloss, 2010), and therefore the distribution of colours across objects in the environment may place an ecological constraint on perception, language and colour categorisation. Adults, infants and non-human animals are able to detect and learn associations between behaviourally relevant objects and their features to guide cognition (Bates et al., 2019; Caves et al., 2018; Hanley et al., 2017; Younger, 1990). Thus humans may learn the distribution of colours in the environment to guide cognition towards the most behaviourally relevant stimuli, resulting in categorical perception of colours.

Recent trends in the literature have focused on the behavioural relevance of objects versus backgrounds for object detection. Patterns in colour statistics and colour naming indicate a warm-cool asymmetry in colour lexicons which supports more efficient communication of behaviourally important warm-coloured objects compared with cool-coloured backgrounds (Gibson et al., 2017; Rosenthal et al., 2018; Twomey et al., 2020; Zaslavsky et al., 2019). However, there is still no clear evidence for a relationship between the colour statistics of the environment and the categorical structure across the full set of BCTs. Although colour is useful for segmenting objects from backgrounds in low-level colour perception (Hansen & Gegenfurtner, 2017), it is also useful for recognising specific objects in high-level colour perception (Bramão et al., 2011). Consequently it was proposed that the origins of BCTs and categorical biases in colour perception may be identified in environmental colour statistics, by studying the information provided by colours about objects for high-level perception.

7.2 Summary of findings

To tackle this problem, the properties of a high-level representation of colour were estimated by testing colour memory. For high-level vision, colours need to be matched between objects, and therefore the dominant noise is due to failure of colour memory (Baddeley & Attewell, 2009; Bae et al., 2015). Under a Bayesian framework, prior knowledge about the world is combined with noisy sensory signals to reduce uncertainty about the sensory signal.

In the first experiment, Bae et al.'s (2015) study was replicated by measuring response frequency for a delayed colour estimation task across a hue circle. Participants were presented with stimuli varying across orange, yellow and green hues for 100ms and, following a 900ms delay, they were asked to estimate the hue. Response frequency was not uniform but rather biased by category structure: responses were biased towards category foci for orange-yellow, green and blue, and away from category boundaries. Thus, a non-uniform prior in colour perception was identified, replicating the results of Bae et al. (2015).

The second study extended the experimental design of Bae et al. (2015) to measure colour memory across hue and saturation for two delay lengths: 900ms and 3900ms. Eighty stimuli were sampled from CIE1931 chromaticity space and displayed on a colour ring. Following the delay, participants browsed an RGB triangle in chromaticity space to select a response. Kernel smoothing was also applied to estimate responses for smaller steps across the colour space. By calculating the divergence of a vector field from target to average response, sinks and sources were identified in the colour space which attracted and repelled responses, and corresponded to Berlin & Kay's (1969) BCTs: green, orange, red, pink, blue and grey. Responses were biased towards more saturated colours (sinks) and away from achromatic colours and category boundaries (sources). The saturation bias had a stronger effect in the long delay condition where uncertainty was greater. As such, categorical biases in a high-level representation of colour were observed across hue and saturation, providing evidence for a categorical prior over colours which directs the visual system towards perceiving and remembering category foci and more saturated colours under conditions of uncertainty.

It was predicted that the non-uniform prior in colour perception originates from the observed distribution of colours in the world, and from distortions to the geometry induced by the task of object recognition. To identify the form of this prior, object colours from ImageNet were randomly sampled, and KDE was used to estimate the probability distribution of colours for a given object. Bayes' rule was applied to calculate the probability of an object given a colour, and variation in probability distributions across the colour space was measured using information theory. Colours with less uniform probability distributions across objects were more informative about object identity. Peaks and troughs in the information geometry did not correspond with BCTs, but there was an effect of saturation: high saturation and zero saturation colours were more informative about objects than low saturation colours, reflecting the average probability of colours in the environment. Highly probable colours were less informative about objects. As such, the model successfully grouped objects into chromatic and achromatic colour categories using the probability of colours in the environment. However, it is not just the colours that are associated with an object which are informative for object recognition, but also the colours of objects that are being discriminated.

Next, a deep neural network was trained to identify objects using the object colours sampled from ImageNet, to test whether complex patterns of information about objects can be learnt across the colour space. The model generated a probability distribution across objects for a given colour. Information theoretic methods were used to analyse how predictions varied across the colour space, and how informative colours are about objects. It was predicted that the model would organise objects into colour categories and, on average, that colour would provide the same amount of information as the entropy of six BCTs. The model organised information provided by colour about objects into five distinct colour categories: red, green, blue, pink and grey, where chromatic colours were most informative about objects, and achromatic colours were least informative about objects. On average, colour did not provide the same amount of information as six colour terms because the colours were not uniformly distributed across the chromaticity space. However, it was observed that distortions to the geometry induced by the object recognition task reflected the non-uniform prior in colour perception, such that colours which were most informative about objects were the colours towards which colour perception was biased. In addition, the colours which were least informative about objects were those which perception is biased away from. These findings support the theory that the colour statistics of objects in the environment may form the basis for a non-uniform prior which directs perceptual processes towards the most informative colours.

For the final study, a deep neural network was trained to identify objects using object colours sampled from Open Images V6, to test whether the distortions to the geometry induced by object recognition for a model trained using ImageNet could be replicated with an alternative dataset. The Open Images model organised information provided by colour about objects into seven distinct colour categories: green, orange/yellow, red, pink, blue, beige and grey. The location of the category foci in the Open Images model closely corresponded with the ImageNet model, and therefore the results were successfully replicated. In addition, as with the ImageNet model, the information geometry of object colour space in the Open Images model followed a similar structure to the geometry of high-level colour space. Since similar colour categorical structures were observed in the distribution of colours across objects from two independent image datasets and in the geometry of high-level colour space, this provides strong evidence that the distribution of colours across objects in the environment, and distortions induced by the task of object recognition, form the basis for a non-uniform prior which directs perceptual processes towards the colours which are most informative about objects. There are statistical patterns of colours in the environment which can explain universal patterns in colour categorisation.

7.3 General conclusions

There are universal constraints in colour categorisation (Berlin & Kay, 1991), and these colour categories impose categorical biases in high-level colour space (Bae et al., 2015). It was suggested that these universal constraints originate from the non-uniform distribution of colours across objects in the environment, and these prior expectations in colour perception reduce uncertainty about the sensory signal. The availability of large and varied image datasets enabled the investigation of how colours are distributed across objects, to test whether human vision is adapted to the features of the environment. As demonstrated here using two independent datasets, there are statistical patterns in the colour properties of objects which do correspond with the geometry of high-level colour space. This provides strong evidence for ecological constraints on colour perception and naming, and demonstrates a relationship between the distribution of colours in the environment and the categorical structure of BCTs. Therefore, it is possible that humans use statistical patterns of colours in the environment as a cognitive prior during perceptual uncertainty, to improve vision and communicate efficiently about colours.

7.4 Evaluation of approach

There were two types of tasks in this project: a colour estimation task and an object recognition task.

7.4.1 Colour estimation task

To extend the colour estimation task into hue and saturation dimensions, a response wheel was replaced by a response triangle. As a result, the stimulus was presented on a ring surrounding the triangle, which allowed the stimulus to be consistently displayed in the same location, rather than the location of the rapid stimulus moving around a large field of view. In addition, the colour of the ring varied as the participant browsed the triangle, so the participants could make an accurate response. This allowed the collection of responses across hue and saturation, but an undelayed condition was not possible as the stimulus could not be displayed during the response. An undelayed task would have provided evidence that the effects observed originate in perception, but there is good evidence of this already (Bae et al., 2015). In addition, the purpose of the present study was to model high-level perception and the effects of memory for the task of object recognition, so an undelayed condition was not required.

A possible adaptation of the colour estimation task to enable measurement of responses without a delay is to display stimuli on one half of the ring (left or right), and collect the response from the other half. An advantage to this design is that it may be possible to identify lateralised effects of categorical perception on hue and saturation. The difference between responses (accuracy, bias or reaction time) to stimuli presented on the left visual field and right visual field could be measured. Findings from previous studies would predict preferential categorical effects on colour perception for stimuli presented in the right visual field (Drivonikou et al., 2007; Franklin, Drivonikou, Bevis, et al., 2008; Gilbert et al., 2006; Roberson et al., 2008; Roberson & Davidoff, 2000; Winawer et al., 2007; Zhou et al., 2010). However, lateralised effects are only observed for very short reaction times (around 400ms: Siuda-Krzywicka et al., 2019), while average response times in the delayed colour estimation tasks here are 3-4 seconds, so it is unlikely that a lateralised effect would be observed.

Another difficulty of extending the task into hue and saturation dimensions is that it introduced edge effects. Hue with constant saturation and luminance varies in a circular distribution in one dimension, so the sampled colour space has no edges. However, saturation varies from low in the centre of the colour space to high on the edge of the colour space. Responses to stimuli on the edge of the triangle are restricted in the direction they can be biased in, so they are biased away from the edge of the triangle and towards its centre. These edge effects cause responses to high saturation stimuli to have larger response errors than colours in the centre. It is not possible to avoid these edge effects as they are simply a feature of multidimensional colour appearance models. Despite these edge effects, it was still possible to identify sinks and sources in the colour space which attracted and repelled responses, and so they were not detrimental to the method, but rather a limitation to the measurement of stimulus-specific response accuracy for a multidimensional colour space.

There was no difference in response accuracy between the two delay conditions, although there was a bias towards more saturated colours in the long delay condition. One reason for not observing a significant difference in response accuracy is the edge effects, which act in the opposite direction to the saturation bias. It is also possible that the difference in delay length between the conditions was not long enough to observe an effect on uncertainty. Olkkonen et al. (2014) also observed that increasing delay length from 0.4s to 4s in a colour estimation task had a significant effect on response bias towards the central colour, but no significant effect on the variability of responses.

An alternative method of increasing uncertainty is by increasing cognitive load, for example by increasing set size or assigning a secondary task during the maintenance interval. Hardman et al. (2017) observed that increasing set size or the difficulty of the secondary task resulted in reduced memory accuracy and more categorical responses for a colour estimation task. Pertzov et al. (2017) and Shin et al. (2017) also observed that set size has a larger effect on memory accuracy than delay duration. As such, future studies may have even more success identifying categorical priors in colour memory by increasing set size rather than delay duration.

A final limitation of the colour estimation study is that all participants were native English speakers. There are universal patterns in colour naming and categorisation, so the participant's responses were used to estimate a universal categorical prior in colour perception. However, there is also cross-cultural variation in the size of colour lexicons, so a language which has the same colour term for green and blue may exhibit slightly different biases in colour perception around the green-blue boundary to those observed here (Roberson et al., 2000). In addition, there are slight differences in the visual environment of different language communities, therefore different cultures may have different communicative needs, resulting in cross-cultural variation in the colours which are important for communication (Twomey et al., 2020). A more accurate estimate of a universal prior would be the average response of participants from a variety of geographic locations and languages. This is more complicated and beyond the scope of this study, but it would be an interesting extension of this work.

7.4.2 Object recognition task

For the object recognition task, the method was limited by the image datasets that were available. These image datasets were used to estimate the properties of the environment, but they may not be entirely representative, as there is a difference between everyday visual experience and web-based content. Firstly, the frequency distribution of object classes within the dataset may differ from the frequency distribution of objects which occur in the environment. For example, the ImageNet dataset contains many object classes which are not common objects, e.g. jinriksha, scabbard, cardoon etc. These objects are not very behaviourally relevant to most people, so most humans are unlikely to recognise them and communicate about them. In addition, neither of the datasets contained many object classes for plants, despite plants being a large component of the visual environment and very behaviourally relevant to humans. Therefore some classes are over or under represented in the dataset with respect to their frequency in the environment.

Furthermore, there may be biases in the list of object classes selected for inclusion in the dataset. The object classes in this project were taken from American English, and thus the objects included are most relevant to North American and Western cultures. If the objects were named in a different language, then the information geometry might reflect the BCTs in that language. There are different food types, materials, flora and fauna which are behaviourally relevant in different environments (Webster et al., 2007), and therefore improvements to the model could be made as more image datasets become available. It might be interesting to compare models trained using images taken in different countries, or objects classified in different languages.

The majority of the object classes in these datasets were man-made. It is possible that man-made objects are designed with colour properties that adhere to the local colour lexicon, as objects would then be easier to identify, remember and communicate about. This could explain the information geometry observed in the object recognition models and would suggest that these colour categories exist in our environment because humans reinforce them. However, this colour distribution is still representative of our environment, and human perception may still be adapted to these statistics, as adaptation is occurs during interaction with the visual environment. It would be interesting to compare models trained to recognise either natural objects or man-made objects to see whether there is a difference in the colours that are the most informative.

The chromaticity of an object's surface was estimated by sampling a random pixel from each image. Chromaticity of an object surface is partially influenced by illumination. Humans have colour constancy, which achieves a robust sensation of surface reflectance despite significant variation in lighting conditions. However, colour constancy requires cues about illumination to resolve any ambiguity about surface chromaticity. By sampling a single pixel from the image, any cues to illumination were removed. The observed clustering of object colours along the Planckian Locus in the ImageNet and Open Images datasets reflects the biases in chromaticity as a result of illumination.

Rather than a single pixel, summary statistics from the image could be used to achieve colour constancy. The distribution of cone contrasts within an objects surface forms a pattern which is roughly constant across illumination changes (Hurlbert et al., 2008; Ling et al., 2008), therefore the mean chromaticity of the object may provide a better estimate of the object chromaticity by reducing noise as a result of illumination (Ling et al., 2008; Milojevic et al., 2018). However, this would require segmentation data to accurately separate the object from background and ensure that only pixels within the objects are included when calculating average chromaticity. While there are algorithms that can segment objects from their background (e.g. snake segmentation in Matlab), this would require a lot of processing power for the size of the dataset, and may also require human verification. In addition, if the object surface has multiple chromaticities, then the mean chromaticity would not be representative of the multimodal chromaticity distribution, and therefore alternative methods such as a mixture model would be required.

Lastly, the object recognition model does not discriminate between objects with variable colour properties and multicoloured objects. For example, a balloon is possible for red, green, blue or yellow, but a lorikeet is red, green, blue and yellow. The model predicts balloons and lorikeets to have the same colour diagnosticity, but lorikeets are more colour diagnostic than balloons to humans, because their colour properties - although very distributed - are highly reliable. For humans, the colour associated with objects in memory is not a single colour but a range of colours (Vurro et al., 2013). Therefore, another limitation of the object recognition model is that it does not represent multicoloured objects as well as single coloured objects, as it is only trained to associate a single colour with each object.

7.5 Extensions to work and future directions

This project has made an additional step towards characterising the colour statistics of the environment, particularly the distribution of hue and saturation across objects. There is strong support that the visual system and linguistic systems are adapted to these distributions. This is useful for the goal of understanding how humans extract useful information about the world to form prior expectations, and there are many more exciting avenues to explore to work towards this goal.

7.5.1 The luminance dimension

Colours were sampled from a 2-D chromaticity space which collapses colours across all luminance levels. As such, in both the colour estimation task and the object recognition task, it is difficult to identify colour terms which only occur at specific luminance levels, such as black, white, grey, yellow and brown.

It would be highly beneficial for future developments to extend the current methods into the luminance dimension. This could be achieved in the colour estimation task by allowing participants to vary the luminance of their response using a sliding bar. Alternatively multiple experiments could be conducted at different luminances, and the results could be aggregated across studies. Kernel smoothing could then be applied to estimate fine grained responses across all three dimensions. This is possible and would be valuable, although it would be very time consuming to collect enough data across all three dimensions of colour space. Another difficulty would be finding participants who are willing to participate in such an extensive study.

Using the luminance of a pixel for object recognition is problematic, as it is strongly influenced by the illumination in the scene, and therefore not very informative about the object's surface reflectance. Humans have lightness constancy to recognise surfaces under different illuminations as the same reflectance. This is achieved by comparing the luminance of surfaces in different parts of the image, and using prior expectations about illumination (Fleming et al., 2003). Lightness constancy is also important for recognising material properties to assist object recognition, such as wet, dry, glossy, metallic and matte (Wiebel et al., 2015).

By sampling a single pixel from each image, it is impossible for the model to learn the proportion of the luminance signal which is the surface reflectance. Attewell and Baddeley (2007) measured the distribution of surface reflectances in four visual environments, which could be used as a prior in Bayesian models of lightness constancy. Alternatively, the surface reflectance of the object can be predicted using image statistics, for example the mean, standard deviation or skewness of the luminance histogram (Motoyoshi et al., 2007; Wiebel et al., 2015). Summary statistics may provide a more precise estimate of the surface reflectance than the luminance of a single pixel by reducing variation in illumination as a result of object shape, but this would require segmentation data for the objects. In addition, objects with multiple surface reflectances need to be considered, for example the mean luminance of a zebra is not representative of the bimodal surface reflectance of its stripy coat. This could be resolved by using a mixture model to estimate the means, or estimating surface reflectance contrast using luminance contrast (Robilotto & Zaidi, 2006). It is possible to take steps towards extending this model into the luminance dimension, but it is not a trivial task.

Extending variation into the luminance dimension in the object recognition model would enable the identification of category foci for black, white and grey in the information geometry. If these achromatic category foci display peaks in information rather than troughs as observed in this project, then this will increase the amount of information provided by colour about objects. While in the present study, the average information provided by colour about objects was less than the entropy of the number of colour categories in the space, including luminance could resolve some of this discrepancy. In addition, it would be interesting to measure whether the colour terms which are earlier in Berlin and Kay's evolutionary sequence (e.g. black, white and red) are more informative about objects than colour terms in later stages.

7.5.2 Behavioural relevance in different contexts

Behavioural relevance of colour categories was operationalised here as the usefulness of colours for object recognition, based on evidence that colours improve object recognition in high-level perception (Bramão et al., 2011). However, there may be other uses of colour categories in high-level perception which also influence the information geometry across colour space. For example, colour is also used for material recognition to discriminate between fine-grained object classes (Adelson, 2001). Colour is more useful for object recognition between structurally similar categories than structurally different categories (Price & Humphreys, 1989), and colour improves human recognition of material changes in objects, for example ripening of fruit, or burning of wood, which is an important cue to the behavioural relevance of the object (Yoonessi & Zaidi, 2010).

It's not just the colours that are most strongly associated with a given object that are important, but also the colours that are *not* associated with a given object, compared with all the possible objects (Schloss et al., 2018). This means that the possible objects in a given scenario will influence which colours are behaviourally relevant, therefore the type of object recognition task will influence the category boundaries which are most important. Some errors in perception are also more costly (Sims et al., 2016), resulting in a higher sensitivity to certain categorical boundaries in risky contexts: a categorical bias in the safest direction would be optimal. For example, a very slight change in facial complexion can be an important indicator of health, so humans are highly sensitive to changes in facial colouration (Stephen et al., 2009). A similar change in the colour of food packaging is not as behaviourally relevant, therefore easily unrecognised.

As such, there are multiple variables within the object recognition task which influence the behavioural relevance of colour. Manipulating these variables, such as the types of object classes or the types of images, may provide interesting insights into the effects of behavioural relevance on categorical biases in colour perception. Category boundaries between BCTs may not be rigid within an individual, but rather they may adapt to improve perception and memory for category boundaries which are important within the context of a particular task.

It would also be interesting to manipulate the behavioural relevance of colours in the colour estimation task and observe the effects on their perceptual representations of colour. Previous research suggests that participants learn to categorise visual features which are behaviourally relevant for a discrimination task (Bates et al., 2019; J. Feldman, 2021; Folstein et al., 2015; Özgen & Davies, 2002). A neural network could be trained to recognise a small group of objects using colour. Human participants could be administered a baseline colour estimation task, followed by training using the same object recognition task, and a test colour estimation task. If categorical biases in colour estimation follow similar patterns as the information geometry in the neural network, then this would suggest that a high-level representation of colour does adapt to behaviourally relevant visual features in the environment.

7.5.3 Cross-cultural research

Both the empirical and computational models here were used to predict BCTs, despite no explicit requirement of colour categorisation in either the colour estimation or object recognition task. Therefore, this method can be applied to cross-linguistic research without the biases induced by traditional colour naming tasks. The colour estimation task is not complex, so the instructions can easily be translated. As more datasets become available, it will also become possible to investigate object classes in different languages, or images from particular environments.

Currently the cross-cultural research on colour naming is dominated by studies which use the Munsell colour chips because the WCS (Cook et al., 2005) used Munsell chips, and this provides the most accessible cross-cultural data on colour categorisation. There are issues not only with the distribution of saturation across Munsell chips used to collect the data, but also with the strict, anglicised criteria for determining the BCTs in a foreign language. It would be a big step in the literature if cross-cultural data could be collected without the biases in the original study, for example the colour estimation task.

7.5.4 Temporal changes in colour perception and language

In this digital age, humans are exposed from infancy to virtual visual environments that are systematically different to their physical environment. If human vision and language is adapted to the environment, then digitisation could have an impact on colour perception and colour lexicons, as the colours which are behaviourally relevant are not only dependent on the immediate physical environment.

Additionally, the prevalence of non-native speakers within a population, along with the size of the language population, can influence language by favouring a less complex and more learnable language (Bentz & Berdicevskis, 2016; Lupyan & Dale, 2016). The Berlin and Kay evolutionary sequence proposes that colour lexicons increase in their accuracy as they evolve by adopting more colour terms to communicate about the same colours. However, the world becomes more globalised, it’s possible that languages will not increase in accuracy but rather reduce in complexity to enable efficient communication among speakers of large populations and across populations in different locations. This highlights the importance of replicating colour naming data such as the WCS (Cook et al., 2005a), as the world has changed so rapidly in the last 20 years, it is possible that this has had an effect on colour perception and colour naming.

7.5.5 Beyond colour perception

The purpose of this project was to estimate non-uniform priors in colour perception using statistics of colours in the environment. This Bayesian model of perception can also be applied to study perceptual biases in other continuous domains, such as orientation (Girshick et al., 2011), shape (Bates et al., 2019; J. Feldman, 2021), motion (Series & Seitz, 2013; Stocker & Simoncelli, 2006; Welchman et al., 2008), time (Jazayeri & Shadlen, 2010), and speech sounds (N. H. Feldman et al., 2009; Holt, 2005). In particular, the perceptual magnet effect has been observed for speech sounds, where, similar to colours, there is reduced discriminability for sounds near phonetic category foci (N. H. Feldman et al., 2009). These non-uniform percepts were successfully modelled using Bayesian methods, whereby listeners use prior expectations about categories to make optimal inferences during perceptual uncertainty.

By using information theory to identify the most informative sensory features for a given task, and comparing them to observed categorical biases in perception, Bayesian theories of perception can be further developed beyond “just-so” stories (Bowers & Davis, 2012).

7.5.6 Overall conclusion

The evidence outlined in this project provides an important step towards demonstrating ecological constraints on colour perception and categorisation, which can explain universal patterns in colour naming. Humans exhibit categorical biases in colour perception towards the colour categories in their language, and these biases can be estimated from the statistical regularities of colours across objects in the environment. This supports a universal mechanism through which sensory systems adapt to environmental statistics, allowing for optimal representation and perception of sensory features, despite limited cognitive resources.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., … Zheng, X. (2016). TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. *ArXiv:1603.04467 [Cs]*. http://arxiv.org/abs/1603.04467

Adelson, E. H. (2001). On seeing stuff: The perception of materials by humans and machines. *SPIE*, *4299*, 1–12. https://doi.org/10.1117/12.429489

Adeyefa-Olasupo, I.-E., & Flombaum, J. I. (2018). *Tutorial Review: Evidence for the Memory Color Effect, 1923–2016*. PsyArXiv. https://doi.org/10.31234/osf.io/ahqsm

Allred, S. R., & Olkkonen, M. (2015). The effect of memory and context changes on color matches to real objects. *Attention, Perception, & Psychophysics*, *77*(5), 1608–1624. https://doi.org/10.3758/s13414-014-0810-4

Ashourian, P., & Loewenstein, Y. (2011). Bayesian Inference Underlies the Contraction Bias in Delayed Comparison Tasks. *PLOS ONE*, *6*(5), e19551. https://doi.org/10.1371/journal.pone.0019551

Attewell, D., & Baddeley, R. J. (2007). The distribution of reflectances within the visual environment. *Vision Research*, *47*(4), 548–554. https://doi.org/10.1016/j.visres.2006.11.015

Baddeley, R., & Attewell, D. (2009). The Relationship Between Language and the Environment: Information Theory Shows Why We Have Only Three Lightness Terms. *Psychological Science*, *20*(9), 1100–1107. https://doi.org/10.1111/j.1467-9280.2009.02412.x

Bae, G.-Y., Olkkonen, M., Allred, S. R., & Flombaum, J. I. (2015). Why some colors appear more memorable than others: A model combining categories and particulars in color working memory. *Journal of Experimental Psychology: General*, *144*(4), 744–763. https://doi.org/10.1037/xge0000076

Baronchelli, A., Gong, T., Puglisi, A., & Loreto, V. (2010). Modeling the emergence of universality in color naming patterns. *Proceedings of the National Academy of Sciences*, *107*(6), 2403–2407. https://doi.org/10.1073/pnas.0908533107

Bates, C. J., Lerch, R. A., Sims, C. R., & Jacobs, R. A. (2019). Adaptive allocation of human visual working memory capacity during statistical and categorical learning. *Journal of Vision*, *19*(2), 11. https://doi.org/10.1167/19.2.11

Belpaeme, T., & Bleys, J. (2009). The Impact of Statistical Distributions of Colours on Colour Category Acquisition. *Journal of Cognitive Science*, *10*(1), 1–20. https://doi.org/10.17791/JCS.2009.10.1.1

Bentz, C., & Berdicevskis, A. (2016). Learning pressures reduce morphological complexity: Linking corpus, computational and experimental evidence. *Proceedings of the Workshop on Computational Linguistics for Linguistic Complexity*, 222–232. http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-309553

Berlin, B., & Kay, P. (1991). *Basic Color Terms: Their Universality and Evolution*. University of California Press.

Biederman, I. (1987). Recognition-by-Components: A Theory of Human Image Understanding. *Psychological Review*, *94*(2), 115.

Biggam, C. P. (2012). *The Semantics of Colour: A Historical Approach*. Cambridge University Press. https://doi.org/10.1017/CBO9781139051491

Bishop, C. M., & Bishop, P. of N. C. C. M. (1995). *Neural Networks for Pattern Recognition*. Clarendon Press.

Bloj, M., Weiß, D., & Gegenfurtner, K. R. (2016). Bias effects of short- and long-term color memory for unique objects. *JOSA A*, *33*(4), 492–500. https://doi.org/10.1364/JOSAA.33.000492

Bosten, J. M., Beer, R. D., & MacLeod, D. I. A. (2015). What is white? *Journal of Vision*, *15*(16), 5. https://doi.org/10.1167/15.16.5

Bowers, J. S., & Davis, C. J. (2012). Bayesian just-so stories in psychology and neuroscience. *Psychological Bulletin*, *138*(3), 389–414. https://doi.org/10.1037/a0026450

Bramão, I., Reis, A., Petersson, K. M., & Faísca, L. (2011). The role of color information on object recognition: A review and meta-analysis. *Acta Psychologica*, *138*(1), 244–253. https://doi.org/10.1016/j.actpsy.2011.06.010

Caves, E. M., Green, P. A., Zipple, M. N., Peters, S., Johnsen, S., & Nowicki, S. (2018). Categorical perception of colour signals in a songbird. *Nature*, *560*(7718), 365–367. https://doi.org/10.1038/s41586-018-0377-7

Chaabouni, R., Kharitonov, E., Dupoux, E., & Baroni, M. (2021). Communicating artificial neural networks develop efficient color-naming systems. *Proceedings of the National Academy of Sciences*, *118*(12). https://doi.org/10.1073/pnas.2016569118

Chollet, F. (2015). *Keras: The Python deep learning API*. https://keras.io/

Cibelli, E., Xu, Y., Austerweil, J. L., Griffiths, T. L., & Regier, T. (2016). The Sapir-Whorf Hypothesis and Probabilistic Inference: Evidence from the Domain of Color. *PLOS ONE*, *11*(7), e0158725. https://doi.org/10.1371/journal.pone.0158725

Conway, B. R., Ratnasingam, S., Jara-Ettinger, J., Futrell, R., & Gibson, E. (2020). Communication efficiency of color naming across languages provides a new framework for the evolution of color terms. *Cognition*, *195*, 104086. https://doi.org/10.1016/j.cognition.2019.104086

Cook, R., Kay, P., & Regier, T. (2005). The World Color Survey Database. In *Handbook of Categorization in Cognitive Science* (pp. 223–241). Elsevier Science Ltd. https://doi.org/10.1016/B978-008044612-7/50064-0

Cuthill, I. C., Allen, W. L., Arbuckle, K., Caspers, B., Chaplin, G., Hauber, M. E., Hill, G. E., Jablonski, N. G., Jiggins, C. D., Kelber, A., Mappes, J., Marshall, J., Merrill, R., Osorio, D., Prum, R., Roberts, N. W., Roulin, A., Rowland, H. M., Sherratt, T. N., … Caro, T. (2017). The biology of color. *Science*. https://doi.org/10.1126/science.aan0221

Danilova, M. V., & Mollon, J. D. (2010). Parafoveal color discrimination: A chromaticity locus of enhanced discrimination. *Journal of Vision*, *10*(1), 4. https://doi.org/10.1167/10.1.4

Davidoff, J., Davies, I., & Roberson, D. (1999). Colour categories in a stone-age tribe. *Nature*, *398*(6724), 203–204. https://doi.org/10.1038/18335

Davidoff, J., & Ostergaard, A. (1988). The Role of Colour in Categorial Judgements. *The Quarterly Journal of Experimental Psychology Section A*, *40*(3), 533–544. https://doi.org/10.1080/02724988843000069

de Fez, M. D., Capilla, P., Luque, M. J., Pérez-Carpinell, J., & Pozo, J. C. del. (2001). Asymmetric colour matching: Memory matching versus simultaneous matching. *Color Research & Application*, *26*(6), 458–468. https://doi.org/10.1002/col.1066

De Valois, K. K., & De Valois, R. L. (2001). Color Vision. In N. J. Smelser & P. B. Baltes (Eds.), *International Encyclopedia of the Social & Behavioral Sciences* (pp. 2252–2256). Pergamon. https://doi.org/10.1016/B0-08-043076-7/03498-7

Donkin, C., Nosofsky, R., Gold, J., & Shiffrin, R. (2015). Verbal labeling, gradual decay, and sudden death in visual short-term memory. *Psychonomic Bulletin & Review*, *22*(1), 170–178. https://doi.org/10.3758/s13423-014-0675-5

Drivonikou, G. V., Kay, P., Regier, T., Ivry, R. B., Gilbert, A. L., Franklin, A., & Davies, I. R. L. (2007). Further evidence that Whorfian effects are stronger in the right visual field than the left. *Proceedings of the National Academy of Sciences*, *104*(3), 1097–1102. https://doi.org/10.1073/pnas.0610132104

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

Ennis, R., Schiller, F., Toscani, M., & Gegenfurtner, K. R. (2018). Hyperspectral database of fruits and vegetables. *Journal of the Optical Society of America A*, *35*(4), B256. https://doi.org/10.1364/JOSAA.35.00B256

Feldman, J. (2021). Mutual Information and Categorical Perception. *Psychological Science*, *32*(8), 1298–1310. https://doi.org/10.1177/0956797621996663

Feldman, N. H., Griffiths, T. L., & Morgan, J. L. (2009). The influence of categories on perception: Explaining the perceptual magnet effect as optimal statistical inference. *Psychological Review*, *116*(4), 752–782. https://doi.org/10.1037/a0017196

Fleming, R. W., Dror, R. O., & Adelson, E. H. (2003). Real-world illumination and the perception of surface reflectance properties. *Journal of Vision*, *3*(5), 3. https://doi.org/10.1167/3.5.3

Folstein, J. R., Palmeri, T. J., Van Gulick, A. E., & Gauthier, I. (2015). Category Learning Stretches Neural Representations in Visual Cortex. *Current Directions in*

Psychological Science, *24*(1), 17–23. https://doi.org/10.1177/0963721414550707

Fortescue, M. (2016). The colours of the Arctic. *Amerindia*, *38*, 25–46.

Fougnie, D., & Alvarez, G. A. (2011). Object features fail independently in visual working memory: Evidence for a probabilistic feature-store model. *Journal of Vision*, *11*(12), 3–3. https://doi.org/10.1167/11.12.3

Franklin, A., Clifford, A., Williamson, E., & Davies, I. (2005). Color term knowledge does not affect categorical perception of color in toddlers. *Journal of Experimental Child Psychology*, *90*(2), 114–141. https://doi.org/10.1016/j.jecp.2004.10.001

Franklin, A., & Davies, I. R. L. (2004). New evidence for infant colour categories. *British Journal of Developmental Psychology*, *22*(3), 349–377. https://doi.org/10.1348/0261510041552738

Franklin, A., Drivonikou, G. V., Bevis, L., Davies, I. R. L., Kay, P., & Regier, T. (2008). Categorical perception of color is lateralized to the right hemisphere in infants, but to the left hemisphere in adults. *Proceedings of the National Academy of Sciences*, *105*(9), 3221–3225. https://doi.org/10.1073/pnas.0712286105

Franklin, A., Drivonikou, G. V., Clifford, A., Kay, P., Regier, T., & Davies, I. R. L. (2008). Lateralization of categorical perception of color changes with color term acquisition. *Proceedings of the National Academy of Sciences*, *105*(47), 18221–18225. https://doi.org/10.1073/pnas.0809952105

Fukuzawa, K., Itoh, M., Sasanuma, S., Suzuki, T., Fukusako, Y., & Masui, T. (1988). Internal representations and the conceptual operation of color in pure alexia with color naming defects. *Brain and Language*, *34*(1), 98–126. https://doi.org/10.1016/0093-934X(88)90126-5

Gao, J., & Sutrop, U. (2014). The basic color terms of Mandarin Chinese: A theory-driven experimental study. *Studies in Language*, *38*(2), 335–359. https://doi.org/10.1075/sl.38.2.03gao

Gibson, E., Futrell, R., Jara-Ettinger, J., Mahowald, K., Bergen, L., Ratnasingam, S., Gibson, M., Piantadosi, S. T., & Conway, B. R. (2017). Color naming across languages reflects color use. *Proceedings of the National Academy of Sciences*, *114*(40), 10785–10790. https://doi.org/10.1073/pnas.1619666114

Gibson, E., Futrell, R., Piantadosi, S. P., Dautriche, I., Mahowald, K., Bergen, L., & Levy, R. (2019). How Efficiency Shapes Human Language. *Trends in Cognitive Sciences*, *23*(5), 389–407. https://doi.org/10.1016/j.tics.2019.02.003

Gilbert, A. L., Regier, T., Kay, P., & Ivry, R. B. (2006). Whorf hypothesis is supported in the right visual field but not the left. *Proceedings of the National Academy of Sciences*, *103*(2), 489–494. https://doi.org/10.1073/pnas.0509868103

Girshick, A. R., Landy, M. S., & Simoncelli, E. P. (2011). Cardinal rules: Visual orientation perception reflects knowledge of environmental statistics. *Nature Neuroscience*, *14*(7), 926–932. https://doi.org/10.1038/nn.2831

Glorot, X., & Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks. *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, 249–256. https://proceedings.mlr.press/v9/glorot10a.html

Groh, A. (2016). Culture, Language and Thought: Field Studies on Colour Concepts. *Journal of Cognition and Culture*, *16*(1–2), 83–106. https://doi.org/10.1163/15685373-12342169

Hanley, D., Grim, T., Igic, B., Samaš, P., López, A. V., Shawkey, M. D., & Hauber, M. E. (2017). Egg discrimination along a gradient of natural variation in eggshell coloration. *Proceedings of the Royal Society B: Biological Sciences*, *284*(1848), 20162592. https://doi.org/10.1098/rspb.2016.2592

Hansen, T., & Gegenfurtner, K. R. (2009). Independence of color and luminance edges in natural scenes. *Visual Neuroscience*, *26*(1), 35–49. https://doi.org/10.1017/S0952523808080796

Hansen, T., & Gegenfurtner, K. R. (2013). Higher order color mechanisms: Evidence from noise-masking experiments in cone contrast space. *Journal of Vision*, *13*(1), 26–26. https://doi.org/10.1167/13.1.26

Hansen, T., & Gegenfurtner, K. R. (2017). Color contributes to object-contour perception in natural scenes. *Journal of Vision*, *17*(3), 14-14.

Hardman, K. O., Vergauwe, E., & Ricker, T. J. (2017). Categorical working memory representations are used in delayed estimation of continuous colors. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(1), 30–54. https://doi.org/10.1037/xhp0000290

Haslam, C., Wills, A. J., Haslam, S. A., Kay, J., Baron, R., & McNab, F. (2007). Does maintenance of colour categories rely on language? Evidence to the contrary from a case of semantic dementia. *Brain and Language*, *103*(3), 251–263. https://doi.org/10.1016/j.bandl.2007.08.007

Heider, E. R., & Olivier, D. C. (1972). The structure of the color space in naming and memory for two languages. *Cognitive Psychology*, *3*(2), 337–354. https://doi.org/10.1016/0010-0285(72)90011-4

Hellige, J. B. (2001). *Hemispheric Asymmetry: What's Right and What's Left*. Harvard University Press.

Helmholtz, H. von. (1867). *Handbuch der physiologischen Optik: Mit 213 in den Text eingedruckten Holzschnitten und 11 Tafeln*. Voss.

Hernández-Andrés, J., Romero, J., Nieves, J. L., & Lee, R. L. (2001). Color and spectral analysis of daylight in southern Europe. *JOSA A*, *18*(6), 1325–1335. https://doi.org/10.1364/JOSAA.18.001325

Holt, L. L. (2005). Temporally Nonadjacent Nonlinguistic Sounds Affect Speech Categorization. *Psychological Science*, *16*(4), 305–312. https://doi.org/10.1111/j.0956-7976.2005.01532.x

Hu, Z., Hanley, J. R., Zhang, R., Liu, Q., & Roberson, D. (2014). A conflict-based model of color categorical perception: Evidence from a priming study. *Psychonomic Bulletin & Review*, *21*(5), 1214–1223. https://doi.org/10.3758/s13423-014-0603-8

Hurlbert, A., Vurro, M., & Ling, Y. (2008). Colour constancy of polychromatic surfaces. *Journal of Vision*, *8*(6), 1101. https://doi.org/10.1167/8.6.1101

Jameson, K. A. (2005). Culture and Cognition: What is Universal about the Representation of Color Experience? *Journal of Cognition and Culture*, *5*(3–4), 293–348. https://doi.org/10.1163/1568537054648527

Jameson, K. A., & Alvarado, N. (2003). Differences in color naming and color salience in Vietnamese and English. *Color Research & Application*, *28*(2), 113–138. https://doi.org/10.1002/col.10131

Jameson, K. A., & D'Andrade, R. G. (1997). It's not really red, green, yellow, blue: An inquiry into perceptual color space. In C. L. Hardin & L. Maffi (Eds.), *Color Categories in Thought and Language* (pp. 295–319). Cambridge University Press. https://doi.org/10.1017/CBO9780511519819.014

Jazayeri, M., & Shadlen, M. N. (2010). Temporal context calibrates interval timing. *Nature Neuroscience*, *13*(8), 1020–1026. https://doi.org/10.1038/nn.2590

Johnson, E. N., Hawken, M. J., & Shapley, R. (2001). The spatial transformation of color in the primary visual cortex of the macaque monkey. *Nature Neuroscience*, *4*(4), 409–416. https://doi.org/10.1038/86061

Jones, C. D., Osorio, D., & Baddeley, R. J. (2001). Colour categorization by domestic chicks. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *268*(1481), 2077–2084. https://doi.org/10.1098/rspb.2001.1734

Kay, P., & Kempton, W. (1984). What Is the Sapir-Whorf Hypothesis? *American Anthropologist*, *86*(1), 65–79. https://doi.org/10.1525/aa.1984.86.1.02a00050

Kay, P., & McDaniel, C. K. (1978). The linguistic significance of the meanings of basic color terms. *Language*, *54*(3), 610–646. https://doi.org/10.1353/lan.1978.0035

Kay, P., & Regier, T. (2003). Resolving the question of color naming universals. *Proceedings of the National Academy of Sciences*, *100*(15), 9085–9089. https://doi.org/10.1073/pnas.1532837100

Kimura, A., Wada, Y., Yang, J., Otsuka, Y., Dan, I., Masuda, T., Kanazawa, S., & Yamaguchi, M. K. (2010). Infants' recognition of objects using canonical color. *Journal of Experimental Child Psychology*, *105*(3), 256–263. https://doi.org/10.1016/j.jecp.2009.11.002

Kingdom, F. A. A. (2011). Lightness, brightness and transparency: A quarter century of new ideas, captivating demonstrations and unrelenting controversy. *Vision Research*, *51*(7), 652–673. https://doi.org/10.1016/j.visres.2010.09.012

Kingma, D. P., & Ba, J. (2017). Adam: A Method for Stochastic Optimization. *ArXiv:1412.6980 [Cs]*. http://arxiv.org/abs/1412.6980

Koenderink, J. J. (2010). *Color for the Sciences*. MIT Press.

Kuehni, R. G. (2001). Determination of unique hues using Munsell color chips. *Color Research & Application*, *26*(1), 61–66. https://doi.org/10.1002/1520-6378(200102)26:1<61::AID-COL6>3.0.CO;2-P

Kuznetsova, A., Rom, H., Alldrin, N., Uijlings, J., Krasin, I., Pont-Tuset, J., Kamali, S., Popov, S., Malloci, M., Kolesnikov, A., Duerig, T., & Ferrari, V. (2020). The Open Images Dataset V4: Unified image classification, object detection, and visual relationship detection at scale. *International Journal of Computer Vision*, *128*(7), 1956–1981. https://doi.org/10.1007/s11263-020-01316-z

Lafer-Sousa, R., Liu, Y. O., Lafer-Sousa, L., Wiest, M. C., & Conway, B. R. (2012). Color tuning in alert macaque V1 assessed with fMRI and single-unit recording shows a bias toward daylight colors. *Journal of the Optical Society of America A*, *29*(5), 657. https://doi.org/10.1364/JOSAA.29.000657

Laws, K. R., & Hunter, M. Z. (2006). The impact of colour, spatial resolution, and presentation speed on category naming. *Brain and Cognition*, *62*(2), 89–97. https://doi.org/10.1016/j.bandc.2006.03.002

Lindbloom, B. (2017, July 4). *RGB/XYZ Matrices*. http://www.brucelindbloom.com/

Lindsey, D. T., & Brown, A. M. (2006). Universality of color names. *Proceedings of the National Academy of Sciences*, *103*(44), 16608–16613. https://doi.org/10.1073/pnas.0607708103

Lindsey, D. T., & Brown, A. M. (2009). World Color Survey color naming reveals universal motifs and their within-language diversity. *Proceedings of the National Academy of Sciences*, *106*(47), 19785–19790. https://doi.org/10.1073/pnas.0910981106

Lindsey, D. T., & Brown, A. M. (2014). The color lexicon of American English. *Journal of Vision*, *14*(2), 17–17. https://doi.org/10.1167/14.2.17

Lindsey, D. T., Brown, A. M., Brainard, D. H., & Apicella, C. L. (2015). Hunter-Gatherer Color Naming Provides New Insight into the Evolution of Color Terms. *Current Biology*, *25*(18), 2441–2446. https://doi.org/10.1016/j.cub.2015.08.006

Ling, Y., Vurro, M., & Hurlbert, A. (2008). Surface chromaticity distributions of natural objects under changing illumination. *Conference on Colour in Graphics, Imaging, and Vision*, *2008*(1), 263–267.

Linhares, J. M. M., Pinto, P. D., & Nascimento, S. M. C. (2008). The number of discernible colors in natural scenes. *JOSA A*, *25*(12), 2918–2924. https://doi.org/10.1364/JOSAA.25.002918

Liu, T., Yuan, Z., Sun, J., Wang, J., Zheng, N., Tang, X., & Shum, H.-Y. (2011). Learning to Detect a Salient Object. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *33*(2), 353–367. https://doi.org/10.1109/TPAMI.2010.70

Livingstone, M. S., & Hubel, D. H. (1987). Psychophysical evidence for separate channels for the perception of form, color, movement, and depth. *Journal of Neuroscience*, *7*(11), 3416–3468. https://doi.org/10.1523/JNEUROSCI.07-11-03416.1987

Lupyan, G., & Dale, R. (2016). Why Are There Different Languages? The Role of Adaptation in Linguistic Diversity. *Trends in Cognitive Sciences*, *20*(9), 649–660. https://doi.org/10.1016/j.tics.2016.07.005

Malkoc, G., Kay, P., & Webster, M. A. (2005). Variations in normal color vision. IV. Binary hues and hue scaling. *JOSA A*, *22*(10), 2154–2168. https://doi.org/10.1364/JOSAA.22.002154

Marr, D. (1982). *Vision: A computational approach.* Freeman and Co.

Masland, R. H. (2001). The fundamental plan of the retina. *Nature Neuroscience*, *4*(9), 877–886. https://doi.org/10.1038/nn0901-877

Milojevic, Z., Ennis, R., Toscani, M., & Gegenfurtner, K. R. (2018). Categorizing natural color distributions. *Vision Research*, *151*, 18–30. https://doi.org/10.1016/j.visres.2018.01.008

Mollon, J. (2006). Monge: The Verriest Lecture, Lyon, July 2005. *Visual Neuroscience*, *23*(3–4), 297–309. https://doi.org/10.1017/S0952523806233479

Mollon, J. D. (2009). A neural basis for unique hues? *Current Biology*, *19*(11), R441–R442. https://doi.org/10.1016/j.cub.2009.05.008

Motoyoshi, I., Nishida, S., Sharan, L., & Adelson, E. H. (2007). Image statistics and the perception of surface qualities. *Nature*, *447*(7141), 206–209. https://doi.org/10.1038/nature05724

Nemes, V. A., Parry, N. R. A., & McKeefry, D. J. (2010). A behavioural investigation of human visual short term memory for colour. *Ophthalmic and Physiological Optics*, *30*(5), 594–601. https://doi.org/10.1111/j.1475-1313.2010.00772.x

Olkkonen, M., McCarthy, P. F., & Allred, S. R. (2014). The central tendency bias in color perception: Effects of internal and external noise. *Journal of Vision*, *14*(11), 5–5. https://doi.org/10.1167/14.11.5

Oxbury, J., Oxbury, S., & Humphrey, N. (1969). Varieties of colour anomia. *Brain : A Journal of Neurology*, *92*, 847–860. https://doi.org/10.1093/brain/92.4.847

Özgen, E., & Davies, I. R. L. (2002). Acquisition of categorical color perception: A perceptual learning approach to the linguistic relativity hypothesis. *Journal of Experimental Psychology: General*, *131*(4), 477–493. https://doi.org/10.1037/0096-3445.131.4.477

Palmer, S. E., & Schloss, K. B. (2010). An ecological valence theory of human color preference. *Proceedings of the National Academy of Sciences*, *107*(19), 8877–8882. https://doi.org/10.1073/pnas.0906172107

Panorgias, A., Kulikowski, J. J., Parry, N. R. A., McKeefry, D. J., & Murray, I. J. (2012). Phases of daylight and the stability of color perception in the near peripheral human retina. *Journal of Vision*, *12*(3), 1. https://doi.org/10.1167/12.3.1

Persaud, K., & Hemmer, P. (2014). The Influence of Knowledge and Expectations for Color on Episodic Memory. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *36*(36), 7.

Pertzov, Y., Manohar, S., & Husain, M. (2017). Rapid Forgetting Results From Competition Over Time Between Items in Visual Working Memory. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *43*(4), 528–536. https://doi.org/10.1037/xlm0000328

Pointer, M. R., & Attridge, G. G. (1998). The number of discernible colours. *Color Research & Application*, *23*(1), 52–54. https://doi.org/10.1002/(SICI)1520-6378(199802)23:1<52::AID-COL8>3.0.CO;2-2

Poralla, J., & Neumeyer, C. (2006). Generalization and categorization of spectral colors in goldfish. II. Experiments with two and six training wavelengths. *Journal of Comparative Physiology A*, *192*(5), 469–479. https://doi.org/10.1007/s00359-005-0082-5

Price, C. J., & Humphreys, G. W. (1989). The effects of surface detail on object categorization and naming. *The Quarterly Journal of Experimental Psychology Section A*, *41*(4), 797–828. https://doi.org/10.1080/14640748908402394

Rathore, R., Leggon, Z., Lessard, L., & Schloss, K. B. (2020). Estimating Color-Concept Associations from Image Statistics. *IEEE Transactions on Visualization and Computer Graphics*, *26*(1), 1226–1235. https://doi.org/10.1109/TVCG.2019.2934536

Regan, B. C., Julliot, C., Simmen, B., Viénot, F., Charles–Dominique, P., & Mollon, J. D. (2001). Fruits, foliage and the evolution of primate colour vision. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *356*(1407), 229–283. https://doi.org/10.1098/rstb.2000.0773

Regier, T., & Kay, P. (2009). Language, thought, and color: Whorf was half right. *Trends in Cognitive Sciences*, *13*(10), 439–446. https://doi.org/10.1016/j.tics.2009.07.001

Regier, T., Kay, P., & Cook, R. S. (2005). Focal colors are universal after all. *Proceedings of the National Academy of Sciences*, *102*(23), 8386–8391. https://doi.org/10.1073/pnas.0503281102

Regier, T., Kay, P., & Khetarpal, N. (2007). Color naming reflects optimal partitions of color space. *Proceedings of the National Academy of Sciences*, *104*(4), 1436–1441. https://doi.org/10.1073/pnas.0610341104

Regier, T., Kemp, C., & Kay, P. (2015). Word Meanings across Languages Support Efficient Communication. In B. MacWhinney & W. O'Grady (Eds.), *The Handbook of Language Emergence* (pp. 237–263). John Wiley & Sons, Inc. https://doi.org/10.1002/9781118346136.ch11

Regier, T., & Xu, Y. (2017). The Sapir-Whorf hypothesis and inference under uncertainty. *Wiley Interdisciplinary Reviews: Cognitive Science*, *8*(6), e1440. https://doi.org/10.1002/wcs.1440

Roberson, D., & Davidoff, J. (2000). The categorical perception of colors and facial expressions: The effect of verbal interference. *Memory & Cognition*, *28*(6), 977–986. https://doi.org/10.3758/BF03209345

Roberson, D., Davidoff, J., & Braisby, N. (1999). Similarity and categorisation: Neuropsychological evidence for a dissociation in explicit categorisation tasks. *Cognition*, *71*(1), 1–42. https://doi.org/10.1016/S0010-0277(99)00013-X

Roberson, D., Davies, I., & Davidoff, J. (2000). Color categories are not universal: Replications and new evidence from a stone-age culture. *Journal of Experimental Psychology: General*, *129*(3), 369–398. https://doi.org/10.1037/0096-3445.129.3.369

Roberson, D., Hanley, J. R., & Pak, H. (2009). Thresholds for color discrimination in English and Korean speakers. *Cognition*, *112*(3), 482–487. https://doi.org/10.1016/j.cognition.2009.06.008

Roberson, D., Pak, H., & Hanley, J. R. (2008). Categorical perception of colour in the left and right visual field is verbally mediated: Evidence from Korean. *Cognition*, *107*(2), 752–762. https://doi.org/10.1016/j.cognition.2007.09.001

Robilotto, R., & Zaidi, Q. (2006). Lightness identification of patterned three-dimensional, real objects. *Journal of Vision*, *6*(1), 3. https://doi.org/10.1167/6.1.3

Rosch, E. (1978). Principles of Categorization. In E. Rosch & B. Lloyd (Eds.), *Cognition and Categorisation* (pp. 27–48). Lawrence Erlbaum.

Rosenthal, I., Ratnasingam, S., Haile, T., Eastman, S., Fuller-Deets, J., & Conway, B. R. (2018). Color statistics of objects, and color tuning of object cortex in macaque monkey. *Journal of Vision*, *18*(11), 1. https://doi.org/10.1167/18.11.1

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., & Fei-Fei, L. (2015). ImageNet Large Scale Visual Recognition Challenge. *ArXiv:1409.0575 [Cs]*. http://arxiv.org/abs/1409.0575

Saunders, B. A. C., & van Brakel, J. (1997). Are there nontrivial constraints on colour categorization? *Behavioral and Brain Sciences*, *20*(2), 167–179. https://doi.org/10.1017/S0140525X97001428

Schloss, K. B., Lessard, L., Walmsley, C. S., & Foley, K. (2018). Color inference in visual communication: The meaning of colors in recycling. *Cognitive Research: Principles and Implications*, *3*(1), 5. https://doi.org/10.1186/s41235-018-0090-y

Scott, D. W. (2012). Multivariate Density Estimation and Visualization. In J. E. Gentle, W. K. Härdle, & Y. Mori (Eds.), *Handbook of Computational Statistics: Concepts and Methods* (pp. 549–569). Springer. https://doi.org/10.1007/978-3-642-21551-3_19

Seabold, S., & Perktold, J. (2010). Statsmodels: Econometric and Statistical Modeling with Python. *Proceedings of the 9th Python in Science Conference*, *57*, 92–96. https://doi.org/10.25080/Majora-92bf1922-011

Series, P., & Seitz, A. (2013). Learning what to expect (in visual perception). *Frontiers in Human Neuroscience*, *7*, 668.

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

Shin, H., Zou, Q., & Ma, W. J. (2017). The effects of delay duration on visual working memory for orientation. *Journal of Vision*, *17*(14), 10. https://doi.org/10.1167/17.14.10

Sims, C. R. (2015). The cost of misremembering: Inferring the loss function in visual working memory. *Journal of Vision*, *15*(3), 2–2. https://doi.org/10.1167/15.3.2

Sims, C. R., Ma, Z., Allred, S. R., Lerch, R. A., & Flombaum, J. (2016). Exploring the Cost Function in Color Perception and Memory: An Information-Theoretic Model of

Categorical Effects in Color Matching. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 2273–2278.

Siuda-Krzywicka, K., Boros, M., Bartolomeo, P., & Witzel, C. (2019). The biological bases of colour categorisation: From goldfish to the human brain. *Cortex*. https://doi.org/10.1016/j.cortex.2019.04.010

Skelton, A. E., Catchpole, G., Abbott, J. T., Bosten, J. M., & Franklin, A. (2017). Biological origins of color categorization. *Proceedings of the National Academy of Sciences*, *114*(21), 5545–5550. https://doi.org/10.1073/pnas.1612881114

Smith, T., & Guild, J. (1931). *The C.I.E. colorimetric standards and their use*. *33*(3), 73–134. https://doi.org/10.1088/1475-4878/33/3/301

Souza, A. S., & Skóra, Z. (2017). The interplay of language and visual perception in working memory. *Cognition*, *166*, 277–297. https://doi.org/10.1016/j.cognition.2017.05.038

Stephen, I. D., Coetzee, V., Smith, M. L., & Perrett, D. I. (2009). Skin Blood Perfusion and Oxygenation Colour Affect Perceived Human Health. *PLOS ONE*, *4*(4), e5083. https://doi.org/10.1371/journal.pone.0005083

Stevens, M., Párraga, C. A., Cuthill, I. C., Partridge, J. C., & Troscianko, T. S. (2007). Using digital photography to study animal coloration. *Biological Journal of the Linnean Society*, *90*(2), 211–237. https://doi.org/10.1111/j.1095-8312.2007.00725.x

Stocker, A. A., & Simoncelli, E. P. (2006). Noise characteristics and prior expectations in human visual speed perception. *Nature Neuroscience*, *9*(4), 578–585. https://doi.org/10.1038/nn1669

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. *Vision Research*, *40*(13), 1711–1737. https://doi.org/10.1016/S0042-6989(00)00021-3

Tanaka, J. W., & Presnell, L. M. (1999). Color diagnosticity in object recognition. *Perception & Psychophysics*, *61*(6), 1140–1153. https://doi.org/10.3758/BF03207619

Tanaka, J., Weiskopf, D., & Williams, P. (2001). The role of color in high-level vision. *Trends in Cognitive Sciences*, *5*(5), 211–215. https://doi.org/10.1016/S1364-6613(00)01626-0

Twomey, C. R., Roberts, G., Brainard, D., & Plotkin, J. B. (2020). What we talk about when we talk about color. *BioRxiv*, 2020.09.29.319517. https://doi.org/10.1101/2020.09.29.319517

Vorobyev, M., & Osorio, D. (1998). Receptor noise as a determinant of colour thresholds. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *265*(1394), 351–358. https://doi.org/10.1098/rspb.1998.0302

Vurro, M., Ling, Y., & Hurlbert, A. C. (2013). Memory color of natural familiar objects: Effects of surface texture and 3-D shape. *Journal of Vision*, *13*(7), 20. https://doi.org/10.1167/13.7.20

Webster, M. A. (1996). Human colour perception and its adaptation. *Network: Computation in Neural Systems*, *7*(4), 587–634.

Webster, M. A., Mizokami, Y., & Webster, S. M. (2007). Seasonal variations in the color statistics of natural images. *Network: Computation in Neural Systems*, *18*(3), 213–233. https://doi.org/10.1080/09548980701654405

Webster, M. A., & Mollon, J. D. (1997). Adaptation and the color statistics of natural images. *Vision Research*, *37*(23), 3283–3298. https://doi.org/10.1016/S0042-6989(97)00125-9

Welchman, A. E., Lam, J. M., & Bulthoff, H. H. (2008). Bayesian motion estimation accounts for a surprising bias in 3D vision. *Proceedings of the National Academy of Sciences*, *105*(33), 12087–12092. https://doi.org/10.1073/pnas.0804378105

Whitehead, R. D., Re, D., Xiao, D., Ozakinci, G., & Perrett, D. I. (2012). You Are What You Eat: Within-Subject Increases in Fruit and Vegetable Consumption Confer Beneficial Skin-Color Changes. *PLOS ONE*, *7*(3), e32988. https://doi.org/10.1371/journal.pone.0032988

Whorf, B. L. (1956). *Language, thought, and reality: Selected writings of Benjamin Lee Whorf*. Technology Press of MIT.

Wiebel, C. B., Toscani, M., & Gegenfurtner, K. R. (2015). Statistical correlates of perceived gloss in natural images. *Vision Research*, *115*, 175–187. https://doi.org/10.1016/j.visres.2015.04.010

Wilken, P., & Ma, W. J. (2004). A detection theory account of change detection. *Journal of Vision*, *4*(12), 11–11. https://doi.org/10.1167/4.12.11

Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences*, *104*(19), 7780–7785. https://doi.org/10.1073/pnas.0701644104

Witzel, C. (2016). New Insights Into the Evolution of Color Terms or an Effect of Saturation? *I-Perception*, *7*(5), 2041669516662040. https://doi.org/10.1177/2041669516662040

Witzel, C. (2019). Misconceptions About Colour Categories. *Review of Philosophy and Psychology*, *10*(3), 499–540. https://doi.org/10.1007/s13164-018-0404-5

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), 1–32. https://doi.org/10.1167/15.8.19

Witzel, C., & Franklin, A. (2014). Do focal colors look particularly 'colorful'? *Journal of the Optical Society of America. A, Optics, Image Science, and Vision*, *31*, A365-74. https://doi.org/10.1364/JOSAA.31.00A365

Witzel, C., & Gegenfurtner, K. R. (2013). Categorical sensitivity to color differences. *Journal of Vision*, *13*(7), 1. https://doi.org/10.1167/13.7.1

Witzel, C., & Gegenfurtner, K. R. (2015). Categorical facilitation with equally discriminable colors. *Journal of Vision*, *15*(8), 22–22. https://doi.org/10.1167/15.8.22

Witzel, C., & Gegenfurtner, K. R. (2018). Are red, yellow, green, and blue perceptual categories? *Vision Research*, *151*, 152–163. https://doi.org/10.1016/j.visres.2018.04.002

Witzel, C., MacDonald, L. W., Biggam, C. P., & Paramei, G. V. (2018). The role of saturation in colour naming and colour appearance. In *Progress in Colour Studies: Cognition, language and beyond*.

Witzel, C., Maule, J., & Franklin, A. (2019). Red, yellow, green, and blue are not particularly colorful. *Journal of Vision*, *19*(14), 27–27. https://doi.org/10.1167/19.14.27

Wuerger, S. M., Atkinson, P., & Cropper, S. (2005). The cone inputs to the unique-hue mechanisms. *Vision Research*, *45*(25), 3210–3223. https://doi.org/10.1016/j.visres.2005.06.016

Wurm, L. H., Legge, G. E., Isenberg, L. M., & Luebker, A. (1993). Color Improves Object Recognition in Normal and Low Vision. *Journal of Experimental Psychology: Human Perception and Performance*, *19*(4), 899.

Yang, J., Kanazawa, S., Yamaguchi, M. K., & Kuriki, I. (2016). Cortical response to categorical color perception in infants investigated by near-infrared spectroscopy. *Proceedings of the National Academy of Sciences*, *113*(9), 2370–2375. https://doi.org/10.1073/pnas.1512044113

Yendrikhovskij, S. N. (2001). A computational model of colour categorization. *Color Research & Application*, *26*(S1), S235–S238. https://doi.org/10.1002/1520-6378(2001)26:1+<::AID-COL50>3.0.CO;2-O

Yoonessi, A., & Zaidi, Q. (2010). The role of color in recognizing material changes. *Ophthalmic and Physiological Optics*, *30*(5), 626–631. https://doi.org/10.1111/j.1475-1313.2010.00722.x

Younger, B. (1990). Infants' Detection of Correlations among Feature Categories. *Child Development*, *61*(3), 614–620. https://doi.org/10.2307/1130948

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

Zaslavsky, N., Kemp, C., Tishby, N., & Regier, T. (2019). Color Naming Reflects Both Perceptual Structure and Communicative Need. *Topics in Cognitive Science*, *11*(1), 207–219. https://doi.org/10.1111/tops.12395

Zeki, S., & Marini, L. (1998). Three cortical stages of colour processing in the human brain. *Brain*, *121*(9), 1669–1685. https://doi.org/10.1093/brain/121.9.1669

Zeki, S., Watson, J., Lueck, C., Friston, K., Kennard, C., & Frackowiak, R. (1991). A direct demonstration of functional specialization in human visual cortex. *The Journal of*

Neuroscience, *11*(3), 641–649. https://doi.org/10.1523/JNEUROSCI.11-03-00641.1991

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, *453*(7192), 233–235. https://doi.org/10.1038/nature06860

Zhang, W., & Luck, S. J. (2011). The Number and Quality of Representations in Working Memory. *Psychological Science*, *22*(11), 1434–1441. https://doi.org/10.1177/0956797611417006

Zhou, K., Mo, L., Kay, P., Kwok, V. P. Y., Ip, T. N. M., & Tan, L. H. (2010). Newly trained lexical categories produce lateralized categorical perception of color. *Proceedings of the National Academy of Sciences*, *107*(22), 9974–9978. https://doi.org/10.1073/pnas.1005669107

Appendices

Appendix A.

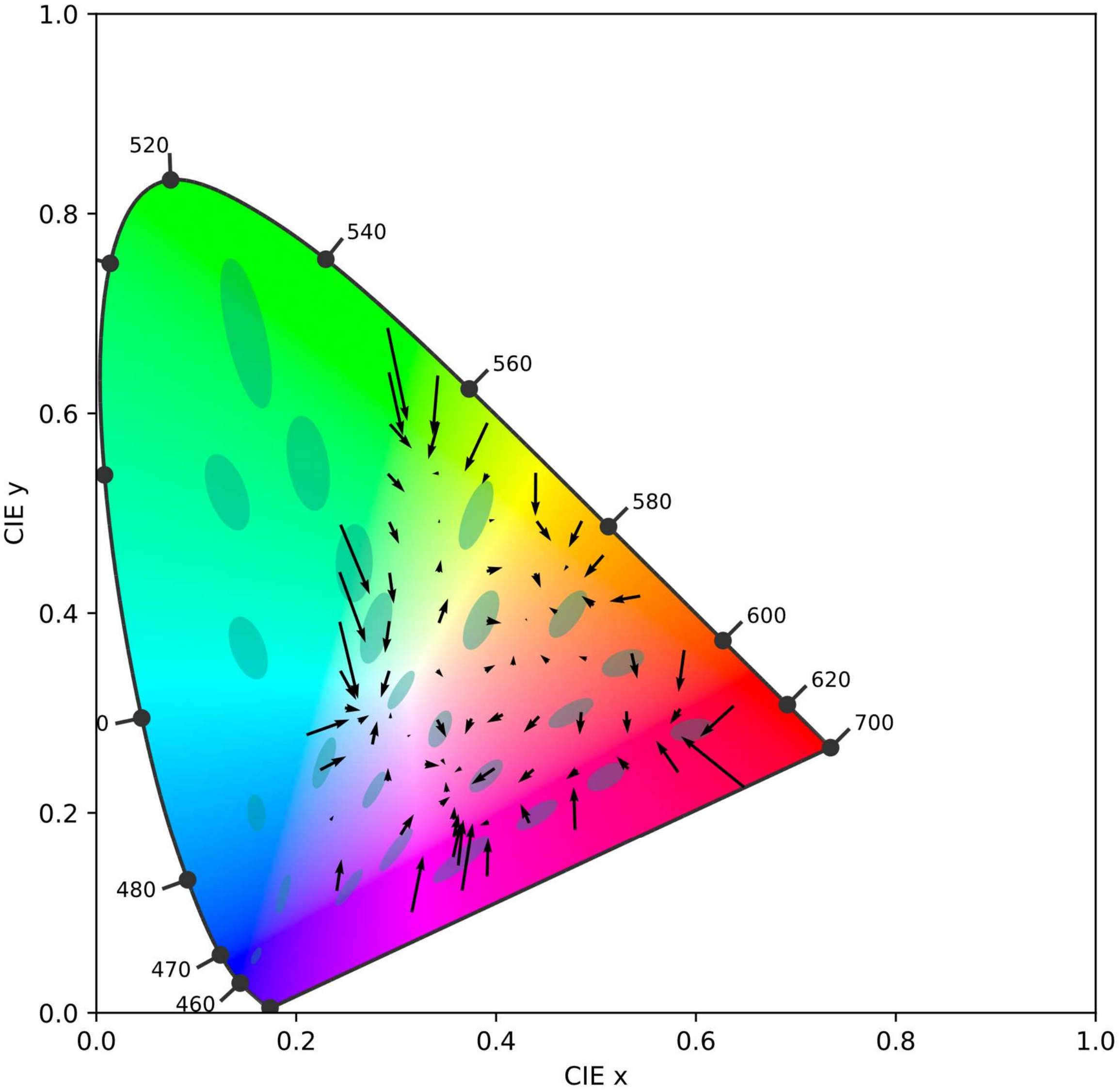


Figure A.1 *MacAdam ellipses on the CIE1931 chromaticity diagram with average response vectors for 900ms delayed colour estimation task.*

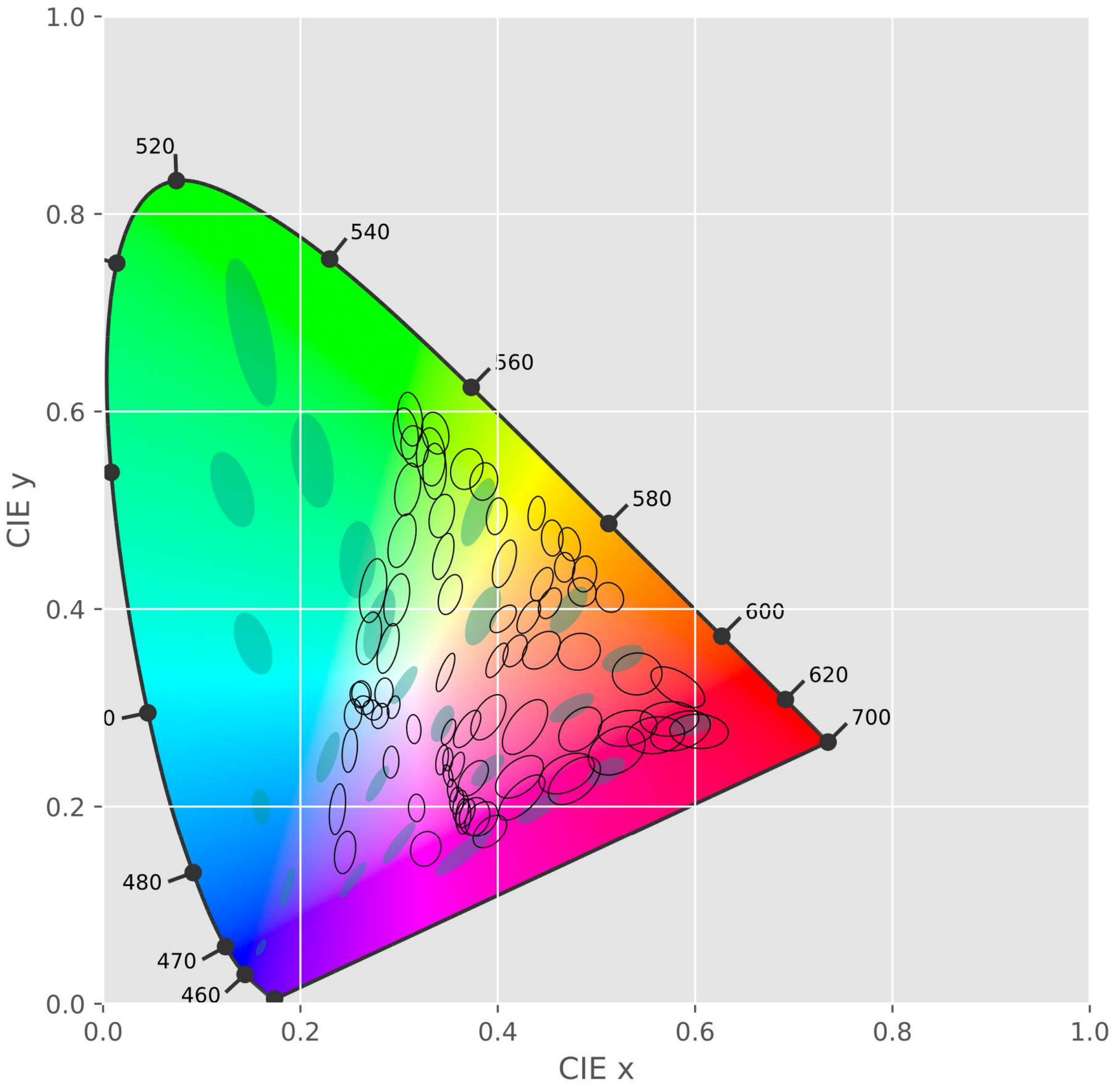


Figure A.2 *MacAdam ellipses on the CIE1931 chromaticity diagram with covariance ellipses for responses in the 900ms delayed colour estimation task.*

Appendix B.

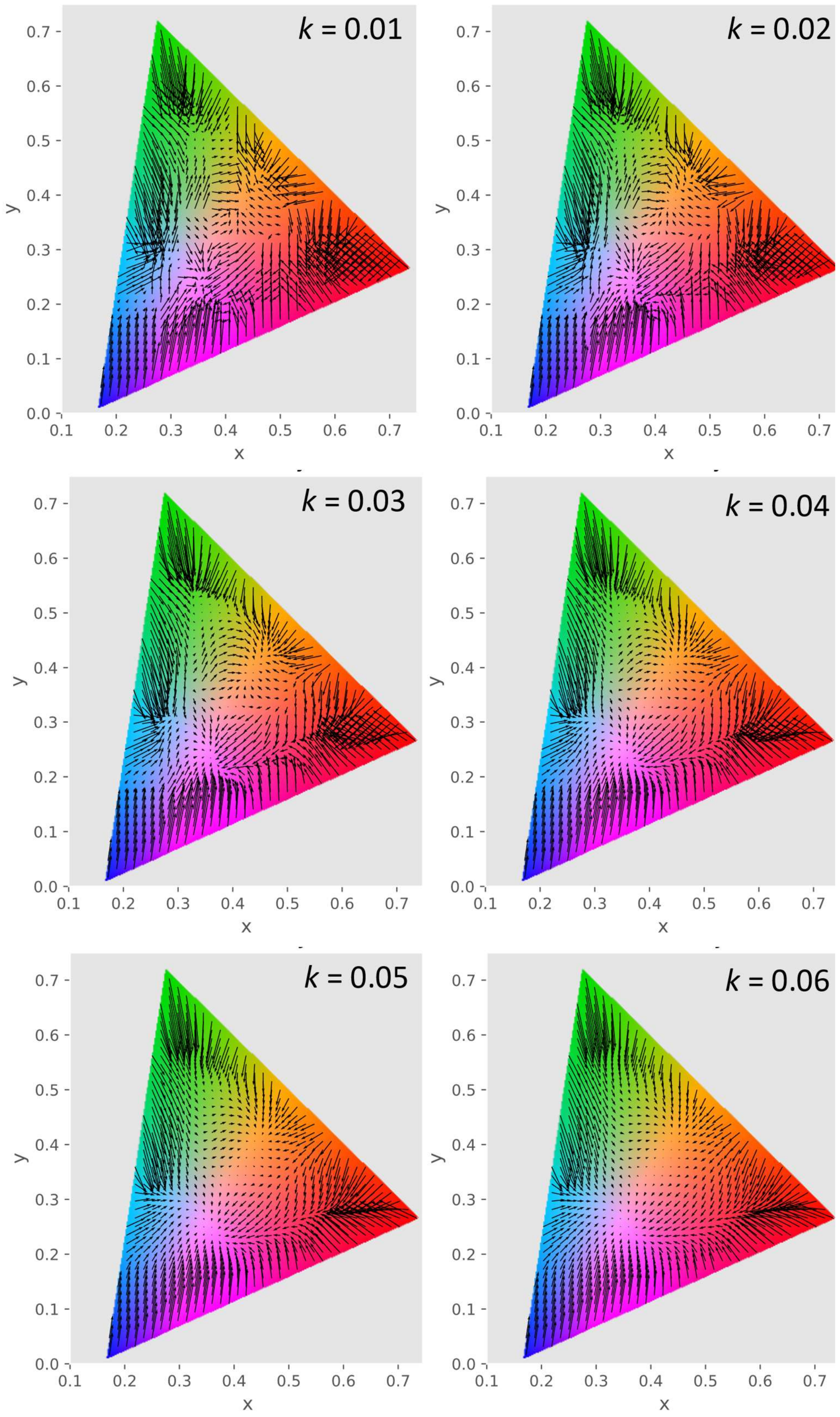


Figure B.1 *KDE for responses to a 900ms delayed colour estimation task for kernel sizes (k) ranging from 0.01 to 0.06.*

Appendix C.

As an alternative to random sampling, a *k*-means clustering algorithm was applied to each image in ImageNet to group the RGB values into clusters. The algorithm iteratively assigns RGB values to one of *k* groups based on the similarity in RGB space, and outputs the centroids of *k* clusters. In this study, *k*=11 clusters were used, such that colours could be clustered into a maximum of 11 BCTs. Only pixels in the centre third of each image were included to reduce the possibility of background colours in the data.

For each image, the centroid of the most dominant cluster was extracted, defined as the cluster with the most values assigned to it. The RGB value of the centroid of the most dominant cluster in each image was used to represent the colour of the object. This RGB value was then converted into chromaticity coordinates using the methods outlined in Section 4.2.

Figure C.1 *K-mean clustering applied to an image of a flower. Pixels in the image were assigned to 11 clusters. Colours on the right are centroids of the 11 clusters, from most dominant to least dominant.*

This *k*-means dataset was used to train a KDE model (using methods outlined in Section 4.3) and an ensemble model of 30 deep neural networks (using methods outlined in Section 5.2).

Mutual information between objects and colours was higher for the KDE model trained on *k*-means sampled data (Table.1) compared to the randomly sampled data (

Table 4.1). However, in general the results are very similar, with a similar information geometry to the model trained on randomly sampled data (Figure 1). As such, for the KDE model, k-means clustering provided no additional insights into the distribution of colours across objects over random sampling.

For the deep neural network, it was observed that the model trained using k-means sampled data did perform more accurately and learnt more mutual information between colours and objects. However, the model was unable to identify category foci and boundaries in the information geometry. This is possibly because k-means clustering did not sample a distribution of colours for each object, but rather just the most dominant colour. Therefore for multicoloured objects, the colours which covered a smaller surface area were not sampled. As a result, complex information about object colour was not learnt from this data. Therefore the k-means sampling method was not further investigated.

Table C.1 *Mutual information between object classes and colours for two models trained on k-means ImageNet dataset. Table shows results for both a KDE model and an ensemble neural network.*

	KDE Model	Neural network (1.62% accuracy)
Uncertainty about colour, H(colour)	11.60 bits	11.60 bits
Uncertainty about colour given object class, H(colour|object)	8.66 bits	10.86 bits
Average amount of information provided by objects about colour, I(colour|object)	11.60-8.66 = 2.94 bits	11.60-10.86 = 0.74 bits
Uncertainty about object class, H(object)	9.79 bits	9.79 bits
Uncertainty about object class given colour, H(object|colour)	7.97 bits	8.22 bits
Average amount of information provided by colour about objects, I(object,colour)	9.79-7.97 = 1.82 bits	9.79-8.22 = 1.57 bits
Average number of colour categories used to discriminate between objects	$2^{1.82}$ = 3.53	$2^{1.57}$ = 2.97

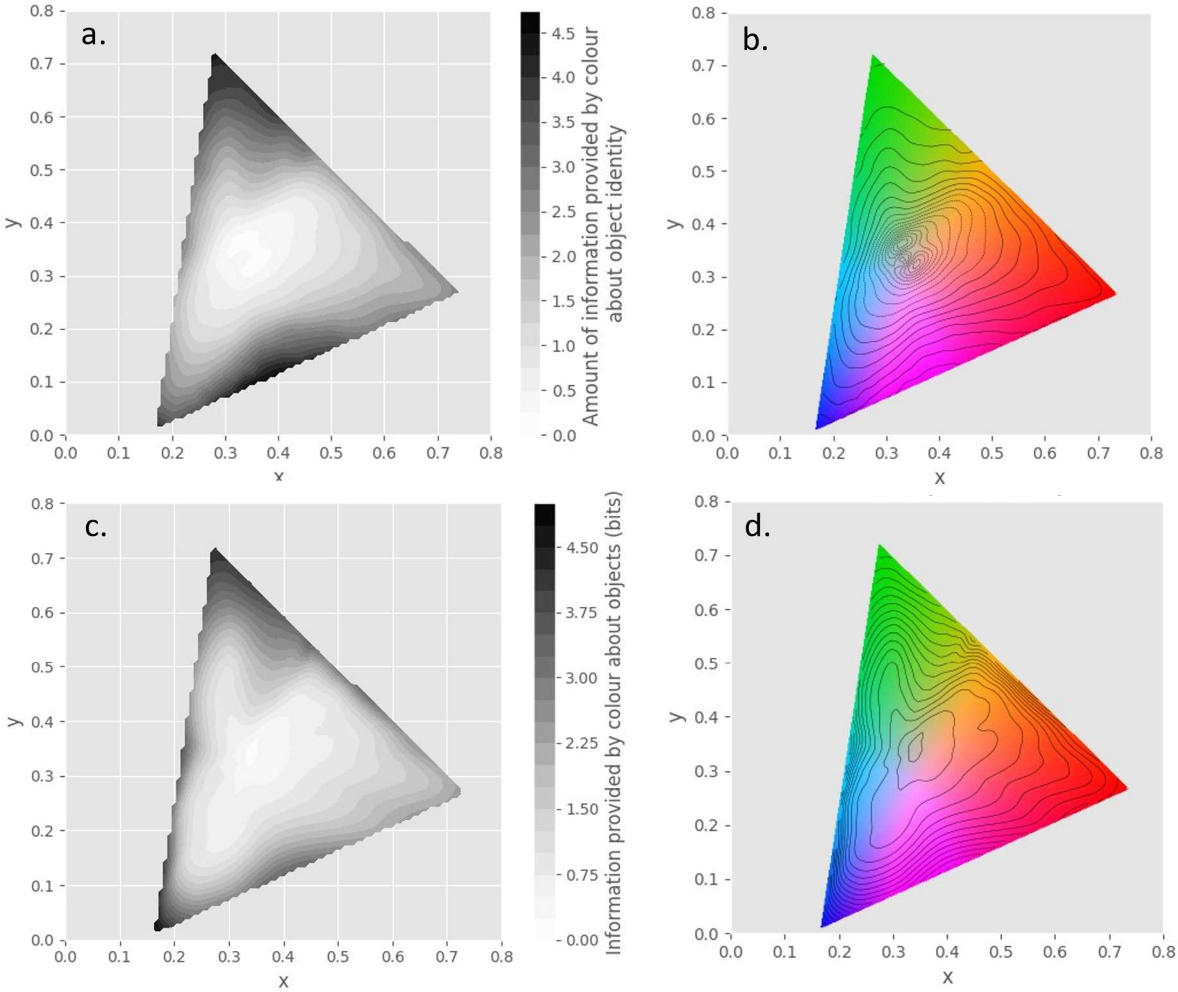


Figure C.2 *Information provided by colour about objects for two models trained on trained on k-means ImageNet dataset. (a-b) Results for KDE model where (a) shows the scale of information and (b) shows the relationship with CIE1931 chromaticity space. (c-d) Results for the ensemble neural network where (c) shows the scale of information and (d) shows the relationship with CIE1931 chromaticity space.*