The Impact of Statistical Distributions of Colours on Colour Category Acquisition*

Tony Belpaeme[1] and Joris Bleys[2]

[1]*University of Plymouth.*
[2]*Vrije Universiteit Brusse*

Different accounts have been proposed to explain the remarkable cross-cultural similarities of human colour categories. One of these accounts, known as empiricism, places the burden of the explanation on the statistical distribution of colours in the shared environment we live in. It has been claimed that this factor is essential for the nature of human colour categories and that it is even capable to account for the evolution of colour categories as observed in human cultures. We present a computational model to verify this hypothesis, which improves on previous work by Yendrikhovskij (2001a). Our results suggest that the chromatic environment does not fully explain the nature of human colour categories, however the distribution of colours in the environment does bias the acquisition of colour categories. We suggest how culture and specifically language might account for the observed nature of human colour categories.

Keywords: colour categorization, computational modelling, chromatic environment, category acquisition

1. Introduction

Even though humans can distinguish an infinite number of colours, we all cut up our colour sensation into a limited number of basic colour categories.

Journal of Cognitive Science 10: 1 - 20, 2009.

* Acknowledgements: We would like to thank Sergej Yendrikhovskij for the fruitful discussions and the data sets he made available. This work is partly funded by the EPSRC CONCEPT project (EP/G008353/1) and by the EU FP7 Integrated Project ITALK. Joris Bleys is funded through a fellowship of the Institute of the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen).

These categories serve many purposes, such as simplifying sensory input or allowing to communicate on what we perceive based on colour information either through discrete labels, i.e. colour terms, or through more complex utterances in which for example basic categories are combined. Ever since Berlin and Kay (1969) and others (Rosch-Heider & Olivier, 1972; Kay & Regier, 2003; Kay et al., 2003) observed that colour categories show a remarkable cross-cultural similarity (although the cross-cultural similarity is still not the accepted position in the field, see Roberson, 2005), there has been an ongoing debate on what the main cause is of this universal character of colour categories. A range of possible hypotheses have been proposed, each highlighting different aspects. Jameson and D'Andrade (1997) and Regier et al. (2007) argue that the similarities can be explained based on some general categorisation principles, such as categories occupying a connected region in the perceptual space and maintaining a maximum dissimilarity between the stimuli that are classified under different colour categories. Some properties of the perceptual colour space restrict the number of possible colour category systems. Komarova and Jameson (2008), using computational models, show how the presence of a small number of colour deficient individuals in a community has the effect of confining colour category boundaries. Others (Van Wijk, 1959; Shepard, 1992; Yendrikhovskij, 2001a) claim this cross-cultural similarity is due to the shared environment in which individuals acquire their colour categories. In this environment, there exists a statistical distribution which captures the presence of different colours in the environment, which is not uniform. It is claimed that this distribution limits the number of possible colour systems. In this work, we wish to replicate and extend the experiments of Yendrikhovskij (2001b, a). Yendrikhovskij demonstrated how the distribution of colours in natural images can be used to extract colour categories that resemble human colour categories. For this purpose, the colour information of 10,000 pixels which were drawn from images of natural scenes was converted to a perceptual colour space and an unsupervised clustering algorithm was used to extract a number of clusters. Yendrikhovskij showed how these clusters resemble the colour categories of American subjects (as determined by Boynton and Olson (1987)). Furthermore, it seemed that when the number of extracted clusters was varied the "evolutionary order" of colour categories (Berlin and Kay, 1969; Kay et al., 1991) could be reconstructed. For

$$\begin{bmatrix}\text{white}\\ \text{black}\end{bmatrix} < [\text{red}] < \begin{bmatrix}\text{green}\\ \text{yellow}\end{bmatrix} < [\text{blue}] < [\text{brown}] < \begin{bmatrix}\text{purple}\\ \text{pink}\\ \text{orange}\\ \text{grey}\end{bmatrix}$$

reference the evolutionary order is shown below.

This theory predicts that cultures having only two colour categories will have a category for dark-cool and one for light-warm colours. A culture with three categories would have an additional category for red. A fourth category will be either green or yellow. This order also excludes some possible colour systems, as before a culture can move on to the next stage, all categories of the previous stages need to be present. This theory predicts that it would be impossible to find a culture which has a category for yellow without having a category for red.

2. Reassessing Yendrikhovskij's Experiments

Yendrikhovskij (2001a, b) (Y) proceeded by taking digital images of natural scenes and sampling 10,000 RGB colour triplets. These triplets were then converted to CIE $L^*u^*v^*$. CIE $L^*u^*v^*$ is a colour space based psychological colour appearance model (Fairchild, 1998). In this colour space the L^* represents the lightness dimension and the u^* and v^* dimension represent the yellow-blue and the red-green opponent channels. This colour space is also known to be perceptually uniform, which means that distance between two colours in this colour space represents the psychological similarity perceived by human subjects (Brainard, 2003). The k-means clustering algorithm was used to extract k clusters from the data. The centroids of the resulting clusters were taken as the colour categories extracted from the data.

To be able to compare the data, Y took the study by Boynton and Olson (1987) (B&O) in which colour names were elicited from subjects when shown isolated colour patches. By calculating the centroid of the colour patches associated with each basic colour term — black, white, red, yellow, green, blue, brown, purple, pink, orange and grey — they were able to determine which colour is identified with which term. In Figure 1 both the focal colours of B&O and the centroids of Y are projected on the u^*v^*-plane — as can be seen from the figure the correspondence was not perfect. A match was made between the centroids and the focal colours shown by the lines in the figure connecting centroids with the human colour categories as measured by B&O. These matches served to compute correlations between centroids and human colour categories.

Using the 10 matches between cluster centroids and B&O focal colours, Y computed correlations between each dimension of the CIE $L^*u^*v^*$ colour space, the chroma C^*_{uv} and the hue H_{uv}. The correlations were high, ranging from r = 0.762 for lightness to 0.999 for hue. Y concludes that "these results support the idea that the structure of colour categories originates from the statistical structure of the perceived environment" (Yendrikhovskij, 2001a). By clustering with increasing values for k, he showed that clusters correspond to the evolutionary order proposed by Berlin and Kay (Berlin & Kay, 1969). Red,

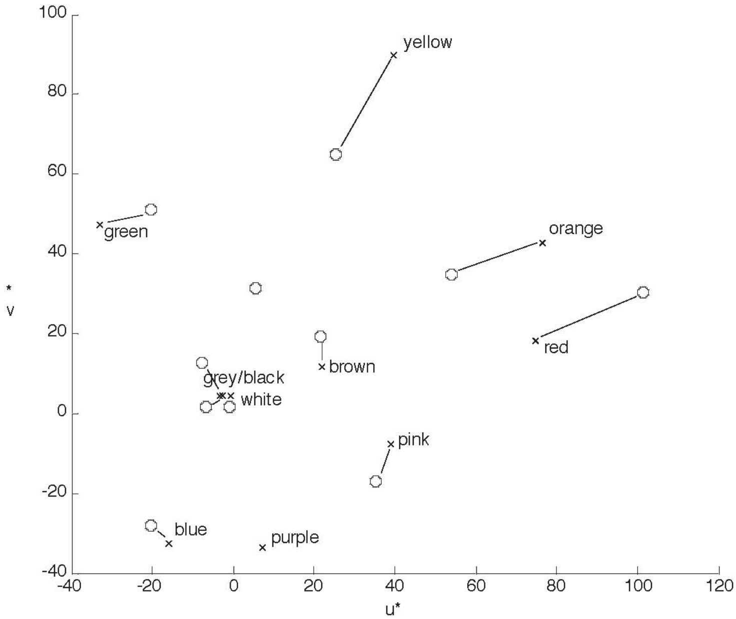

Figure 1. Mapping between colour categories from the computational model (circles, based on Y) and English speakers (crosses, based on B&O) projected on the u^*v^* -plane. No cluster for purple was found. Instead a cluster for green-yellow was found, which can not be linked to any basic colour.

for example, appears quite early when clustering because the colour red is abundant in the data and because the location of red in the perceptual space is furthest from black and white. Green, while abundant in the data, emerges after red because its location is too close to black and white for the cluster algorithm to select it as a cluster. He concludes that "the evolutionary order of colour terms depends on both the external properties of the outside world (frequency of colour occurrence) and the internal properties of the perceptual system (metrics of colour space)" (Yendrikhovskij, 2001a).

Without denying the significance of Yendrikhovskij's work, we would like to critically assess the evidence presented and extend his work to explore the hypothesis that the distribution of colours in the environment plays a crucial role in the nature of human colour categories even further. The following elements of his research pose problems for the conclusions drawn.

1. In order to truly validate the claim that the high correlations to human colour categories are mainly due to the colour distributions in the environment, it is essential to compare the reported results starting from a control dataset in which no such distribution is present (e.g. each colour occurs with the same probability). Only if the correlations between the centroids found in the latter dataset is significantly lower than those originally reported, can one conclude that the high correlation in the original study is due to the colour distribution present in the original dataset. In order to measure the importance of the colour distribution, one could also start from a different set of pictures and compare the results to those reported in the original study.

2. As the perception of chromatic stimuli is unique to every person, caused by, among others, lens brunescence (Lindsey & Brown, 2002) and variation in opsin genes (Winderickx et al., 1992; Neitz et al., 1993; Sharpe et al., 1999), it would be interesting to verify whether the results still hold using different colour appearance models. In the original Y study only one such model (based on CIE $L^*u^*v^*$ colour space and PAL television viewing conditions) was explored.

These comments define an agenda for new experiments:

1. We will collect data sets from different environments, including a control environment in which no statistical information is present. This allows us to test the hypothesis that different ecologies still produce the same division of the

colour continuum and to what degree they correlate to human colour categories. This will allow us to argue whether the environment truly determines human colour categorisation or not.

2. We will analyze the data under two different colour appearance models, CIE $L^*a^*b^*$ and CIE $L^*u^*v^*$, instead of just CIE $L^*u^*v^*$. This allows us to study the impact of the peculiarities of the colour space, i.e. its "nooks and crannies", on the reported results.

3. Reimplementing The Experiments

3.1. Collection of colour stimuli

Two image collections were gathered, one containing nature images and one urban images.[1] The nature collection was compiled from image databases on the internet, and contains imagery of animals, flowering plants and landscapes. The urban collection contains images shot with a digital camera (Olympus C-4000 ZOOM) in Northern European urban environments; it contains imagery of buildings, people and urban activities, both indoor and outdoor. Both collections contain 300 images.

From both image sets 25,000 RGB-pixels were randomly drawn, we will refer to these datasets as NATURAL and URBAN. We also added a control dataset, called RANDOM, which consists of 25,000 random RGB-values; this dataset will be used to test the null-hypothesis that categories are not influenced by the chromatic distribution in the environment. All RGB-values of the datasets were transformed to CIE $L^*a^*b^*$ and CIE $L^*u^*v^*$ using standard formulae (Wyszecki & Stiles, 1982) found in Appendix I.

A first analysis of the data reveals that natural and urban environments indeed contain non-random structure. Figures 2, 3 show histograms of the CIE $L^*a^*b^*$ and CIE $L^*u^*v^*$ values of the datasets. While the RANDOM set has a quasi uniform distribution,[2] the NATURAL and URBAN have a higher

---

[1] The images and data can be obtained from the authors' website at http://arti.vub.ac.be

[2] Unlike the technical RGB representation for colour, the shape of the used colour appearance models is not a cube but rather an irregular convex shape. One might have the impression from figures II and III that the RANDOM dataset is not

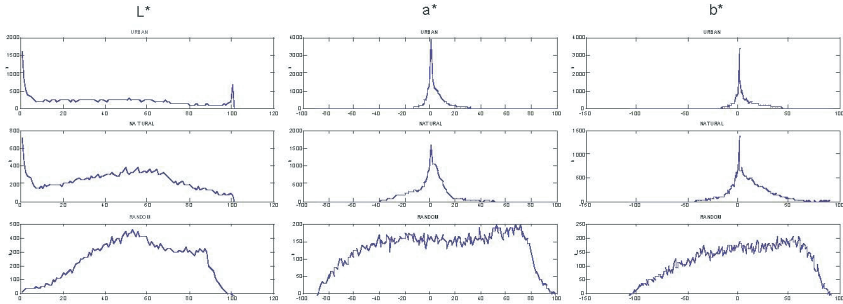

Figure 2. Histogram for CIE L^*, a^* and b^* values of the URBAN (top), NATURAL (middle) and RANDOM (bottom) datasets. The RANDOM dataset has a quasi uniform distribution. The NATURAL and URBAN dataset have a higher concentration of lowly saturated colours.

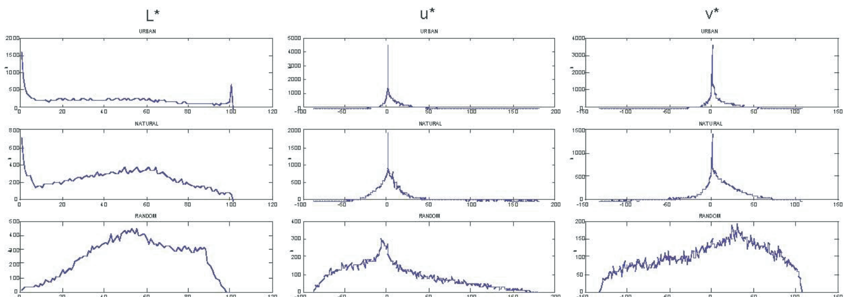

Figure 3. Histogram for CIE L^*, u^* and v^* values of the URBAN (top), NATURAL (middle) and RANDOM (bottom) datasets. The RANDOM dataset has a quasi uniform distribution. As in Figure 2, the NATURAL and URBAN dataset show more lowly saturated colours.

distribution of lowly saturated colours. This reconfirms previous observations on the chromatic content of natural scenes (Hendley and Hecht, 1949; Burton and Moorhead, 1987; Howard and Burnidge, 1994), where it was found that natural occurring colours occupy a restricted area of the chromaticity diagram

---

uniformly distributed, but this is not so.

(Yendrikhovskij, 2001a).

3.2. Extracting colour categories

As clustering algorithm we used the k-means clustering algorithm (Lloyd, 1982). *k*-means clustering uses an iterative re-estimation procedure. Initially, *k* data points are selected as initial centroids. Then all data points are assigned to the nearest centroid (according to the distance between the sample and the centroid) and the centroid is recalculated to be the mean of all the samples that are associated to it after which all samples are classified again. The algorithm continues until a stop criterion is met, usually when there is no further change in the assignment of the data points.

As *k*-means clustering is not deterministic (a random seed is needed to select the initial centroids), the clusters found by each run of the algorithm might vary. How much the clusters vary depends on the structure of the data. To deal with possible variation in the found clusters, the colour data was clustered 1000

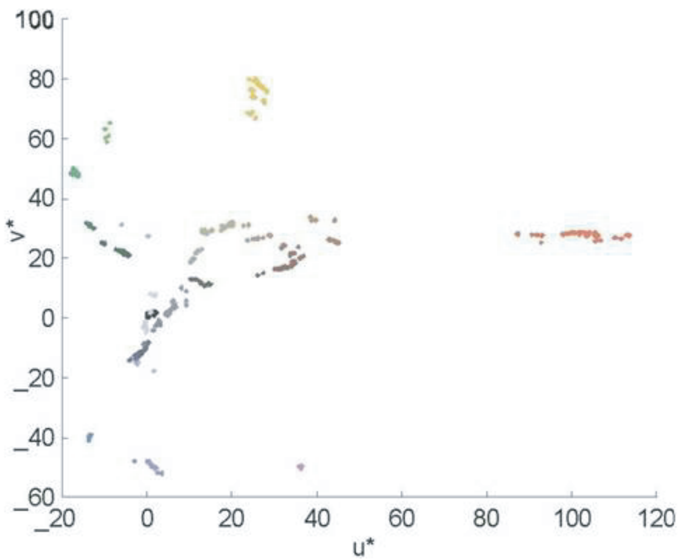

Figure 4. Typical example of the variation of the centroids found after running the standard *k*-means clustering algorithm. The collection of the centroids of 1000 independent runs of the algorithm is projected on the u^*v^* -plane for the NATURE dataset for $k = 11$.

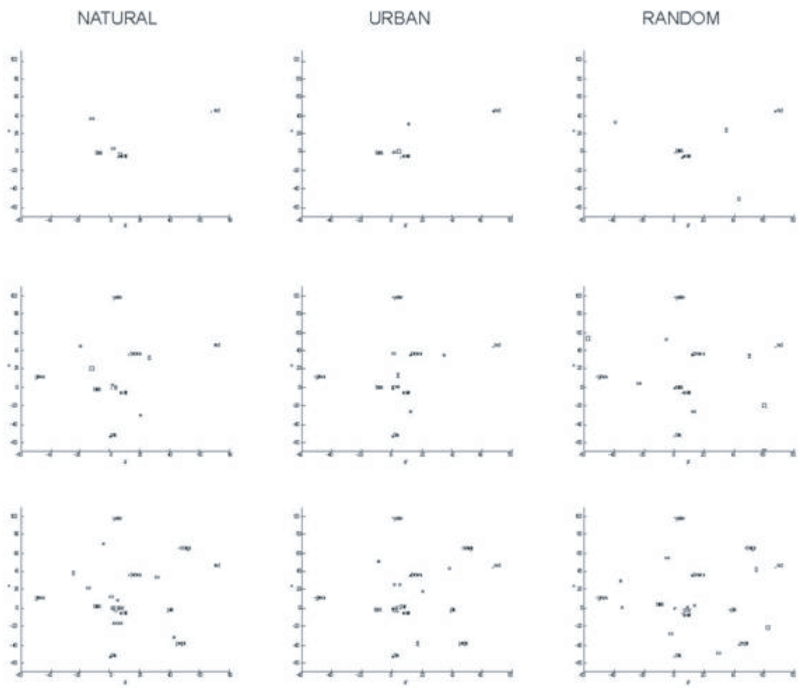

Figure 5. Extracted categories plotted in the a^*b^*-plane (squares). For reference, corresponding human colour categories according to Sturges and Whitfield are included on each plot (dots). The left series show categories extracted from NATURAL data, the middle series from URBAN data and the right series from RANDOM data. For each series three, seven and eleven categories were extracted.

times. The variation in the outcome of the centroids found by the algorithm for the datasets we are using is illustrated in Figure 4. The $1000 \times k$ centroids are then clustered again using k-means clustering, but now the initial points are each time chosen to be equal to the outcome of one run in the previous phase. From these solutions, the one in which the average distance between the $1000 \times k$ centroids and the nearest centroid in the solution is minimal, is chosen to be the final k clusters (see Figure 5). The goal of this additional step is to avoid some algorithm specific problems due to outliers in the original dataset (Bradley and Fayyad, 1998).

3.3. Comparing with human colour categories

In order to evaluate how the clusters compare to human colour categories, we take the study by Sturges and Whitfield (1995) (S&W). Yendrikhovskij used data by Boynton and Olson (Boynton & Olson, 1987) (B&O) to compare his extracted categories with. S&W replicated the B&O experiments and, important for our purpose, included black and white in their studies. They defined the location of eleven basic surface colours (black, white, red, green, yellow, blue, brown, grey, orange, purple and pink) using a monolexemic naming technique. For this, twenty English subjects named a set of 446 colour chips drawn from the Munsell colour set. We use the foci for the 11 colour names, which were converted from Munsell coordinates to CIE $L^*a^*b^*$ or CIE $L^*u^*v^*$ using the Munsell colour conversion tables (Wyszecki & Stiles, 1982).

In order to compare cluster centroids and human colour categories, we first need to match each centroid with a colour category. We do this using an automated technique, which searches the best match between centroids and categories. The best match minimises the sum of Euclidean distances between each centroid-category tuple. Note that this matching is different to that from the original Y paper, as it requires a match between each centroid and each category.

For 7 clusters this match was computed to the 7 colour categories corresponding to the 7 colour names that appear first in the evolution of colour names according to (Berlin and Kay, 1969): black, white, red, green, yellow, blue and brown.[3] For 11 clusters we additionally used the orange, pink, grey and purple foci. Figures 6, 7 show scatter plots of the cluster centroids and human colour categories for the lightness, hue (H) and chroma (C^*) dimensions in CIE $L^*a^*b^*$ and CIE $L^*u^*v^*$. The hue and chroma values can be computed using the equations in Appendix II.

4. Comparison and Analysis

To compare the impact of the statistical distribution on the resulting

[3] This assumption might be hard to fully justify, but we are not aware of any study that tries to reconstruct the exact location of the colour categories in the history of English.

Table 1. Correlation between cluster centroids and human colour categories in the CIE $L^*a^*b^*$ colour space.

7 categories	L^*	a^*	b^*	C^*_{ab}	H_{ab}
Natural	0.781*	0.429	0.714*	0.905*	0.619
Urban	0.781*	0.238*	1.000*	0.524	0.429
Random	0.781*	0.429	0.714*	0.905	0.619

11 categories	L^*	a^*	b^*	C^*_{ab}	H_{ab}
Natural	0.785*	0.200	0.745*	0.709*	0.636*
Urban	0.935*	0.382	0.745*	0.491*	0.345
Random	0.411	0.309	0.782*	0.600*	0.7098*

*Correlation is significant at the 0.05 level.

Table 2. Correlation between cluster centroids and human colour categories in the CIE $L^*a^*b^*$ colour space.

7 categories	L^*	u^*	v^*	C^*_{uv}	H_{uv}
Natural	0.878*	0.619	0.619	1.000*	0.714*
Urban	0.976*	0.619	0.905*	0.714*	0.905*
Random	0.293	0.714*	0.429	0.714*	0.714*

11 categories	L^*	u^*	v^*	C^*_{uv}	H_{uv}
Natural	0.561*	0.673*	0.636*	0.709*	0.745*
Urban	0.972*	0.564*	0.673*	0.636*	0.636*
Random	0.187	0.418	0.745*	0.491*	0.818*

*Correlation is significant at the 0.05 level.

centroids several analyses can be performed. We provide the results of two such analyses.

The first analysis is based on a correlation measure (Kendall's Tau) for each dimension.

This non-parametric test does not require data to have a certain distribution (Conover, 1999). The test returns values between -1 and 1. A value of 1

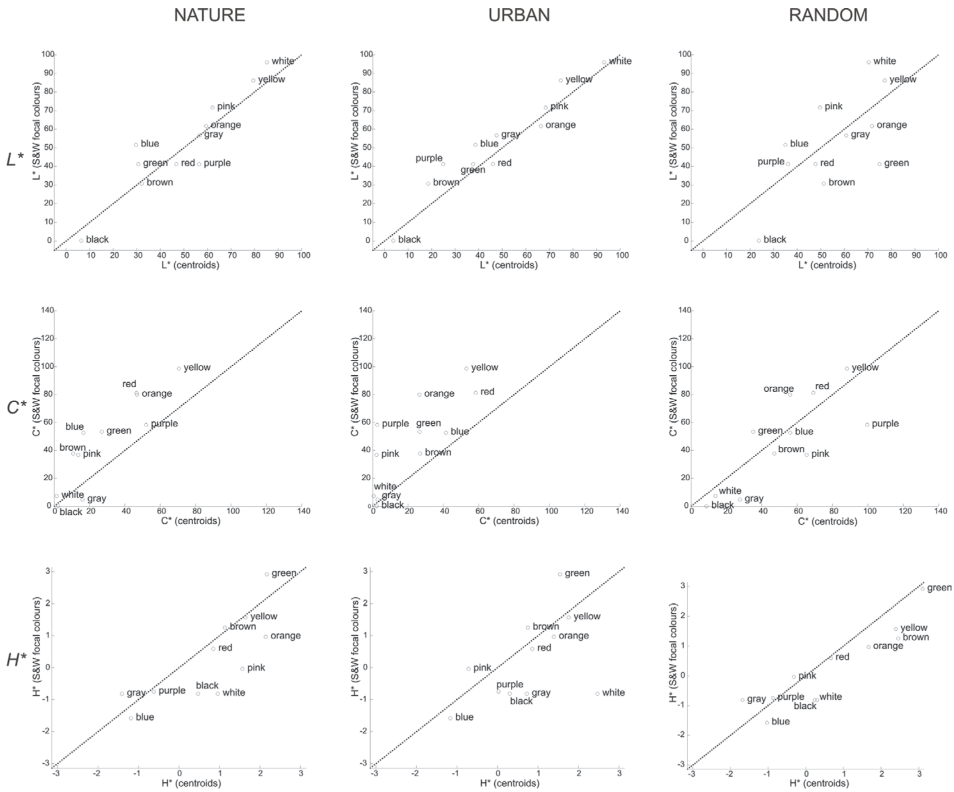

Figure 6. Scatter plots of 11 extracted centroids against 11 focal colours of S&W in CIE $L^*a^*b^*$ dimensions. The plots show respectively the L^*, C^*_{ab} and H_{ab}-dimensions and the centroids extracted from the NATURE, URBAN and RANDOM dataset.

indicates that the correlation is total and -1 that the correlation is total, but inverse. A value of 0 indicates that no linear correlation could be found, while values closer to 1 or -1 indicate an increasing correlation. The correlation was computed for 7 and for 11 clusters. Table I and II report the correlations between the centroids extracted from the NATURAL, URBAN and RANDOM dataset and the S&W categories, both for CIE $L^*a^*b^*$ and CIE $L^*u^*v^*$.

The high correlations between the centroids extracted from the NATURE

Table 3. Average correlations between 11 clusters from the NATURAL, URBAN and RANDOM data set in the CIE $L^*u^*v^*$ and CIE L*a*b* colour spaces.

	Natural-Urban	Urban-Random	Natural-Random
CIE $L^*a^*b^*$	0.5928	0.4618	0.622
CIE $L^*u^*v^*$	0.694	0.491	0.614

and URBAN dataset and the S&W categories confirm the results of Yendrikhovskij. However, the correlation remains high (although somewhat lower) for the RANDOM dataset. As the RANDOM dataset served to test the null-hypothesis, one would expect the correlation to be zero on average, as the RANDOM data set contains no chromatic structure, as opposed to the NATURE or URBAN dataset. Nevertheless, the uniform distribution of the RANDOM set still results in clusters having a remarkable positive correlation with human colour categories. This can only be explained by the structure of the colour space having an influence on the clustering.

Table 3 shows the mean correlations between the clusters found for all three data sets (NATURAL, URBAN and RANDOM). The correlation is in all cases positive. It is noteworthy that the clusters extracted from the RANDOM set correlate well with clusters from the NATURAL and URBAN set. This suggests that the clusters are not so much under the influence of the distribution of the chromatic input, but rather that they are the result of biases introduced by the colour appearance model and the clustering algorithm.

For the second analysis, we created a benchmark test to compare the performance of each resulting set of centroids in more detail. This benchmark consists of naming one hundred[4] consensus chips (chips for which there was unanimous agreement in colour naming) of the study of Sturges and Whitfield. In order to name these chips, they use the colour term that is assigned to the centroid which is closest to the chip (1 nearest neighbour classification) as this is the same classification mechanism that is used during k means clustering. For the centroids that are a result of the algorithm, the terms are copied from

[4] The actual study reports on 102 consensus chips. 2 blue chips were left out for the benchmark due to issues related to the conversion between different colour spaces.

Table 4. Number of correctly named consensus samples broken down by category: white (WE), grey (GY), black (BK), green (GN), yellow (YW), blue (BL), red (RD), purple (PU), brown (BR), orange (OR) and pink (PK). The total number of consensus chips is shown on top, The top part represents the results in CIE $L^*a^*b^*$, the bottom part in CIE $L^*u^*v^*$. The results are shown for the categories found in S&W and three datasets: NATURE (NAT), URBAN (URB) and RANDOM (RAN). Only the results for 11 centroids are shown.

	WE	GY	BK	GN	YW	BL	RD	PU	BR	OR	PK	total
	2	6	3	22	8	25	4	14	4	6	6	100
S&W	2	6	3	17	8	18	4	9	4	6	6	83
NAT	2	1	3	8	8	11	4	6	2	0	0	45
URB	2	4	2	5	8	18	4	2	0	0	3	48
RAN	2	0	3	2	3	5	4	0	0	1	3	20
S&W	2	6	3	18	8	18	4	9	4	6	6	84
NAT	2	5	3	15	8	24	4	6	0	0	1	68
URB	2	5	3	5	8	24	4	6	2	0	2	61
RAN	2	0	3	13	8	0	4	1	0	4	4	39

the category of S&W they matched with. The results, broken down by each category, are summarized in Table 4.

A first important observation from these results, is that even when the categories from

S&W are used, the benchmark only reaches about 83% success. This suggests that the one-nearest neighbour classification algorithm, although capable of accounting for more than three quarters of the consensus chips, might be too simple to capture all the richness of human colour categories.[5] This first results sets the maximal level of success one might hope to achieve when using this particular classification algorithm.

The performance of the centroids resulting from the RANDOM data set is still quite good (a quarter to a half of the maximal expected performance, depending on the used colour space). This can only be accounted for by the

[5] This also points to the idea that focal colours (the chip which subject named with the lowest response time) for colour categories might not be in the centre of all the chips it classifies.

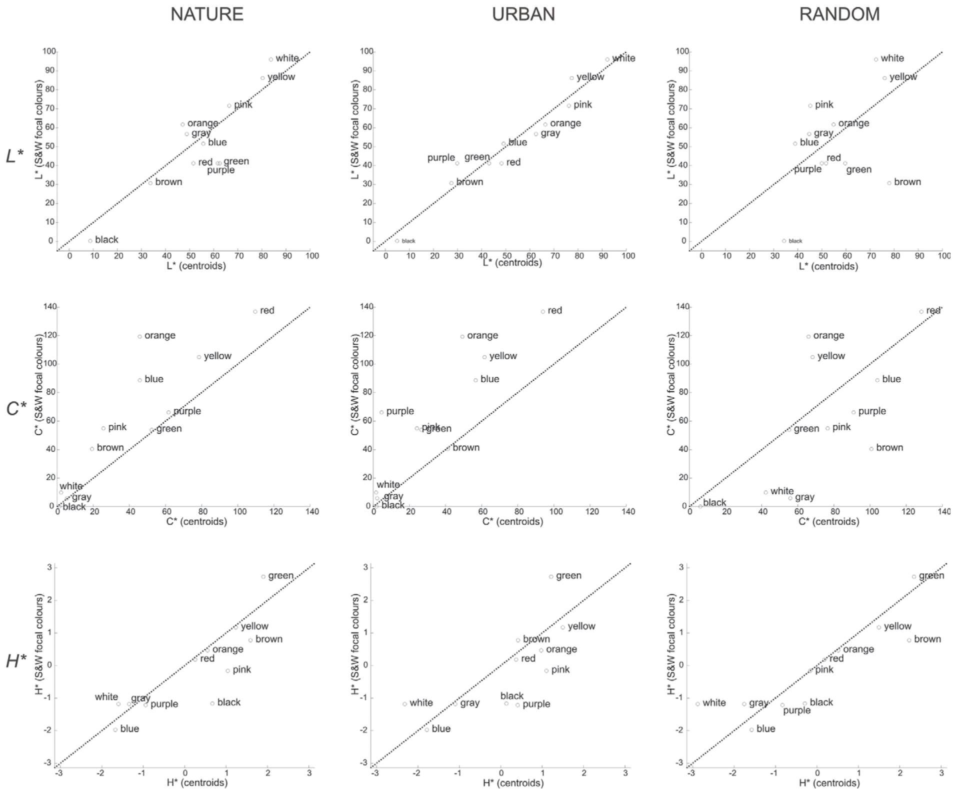

Figure 7. Scatter plots of 11 extracted centroids against 11 focal colours of S&W in CIE L*u*v* dimensions. The plots show respectively the L*, C*uv and Huv - dimensions and the centroids extracted from the NATURE, URBAN and RANDOM dataset. The more a point lies on the diagonal, the better the automatically extracted centroid matches the human colour category.

shape of the colour spaces used and the clustering algorithm that was used but not to any statistical distribution in the environment. However, if such a distribution is present in the data set such as in the NATURE and URBAN data set, it significantly improves the performance of the benchmark by about a third of the maximal expected performance. About half or a quarter of the maximal expected performance, when using the CIE *L*a*b** or CIE *L*u*v** colour space, remains unaccounted for.

5. Discussion

We have presented a model in which categories are automatically extracted from a set of colour images and compared to human colour categories determined by (Sturges andWhitfield, 1995). A correlation measure and a benchmark was used to quantify the influence of the environment on the acquisition of colour categories.

The chromatic structure in the environment has a positive impact on the correlation of the resulting centroids of the clustering algorithm and colour categories of human subjects. However, even if there is no chromatic structure in the environment, our model still manages to extract categories which correlate well with human colour categories. This suggests that while there is an impact of the environment on resulting colour categories, the perceptual colour space has a more profound influence.

The achieved correlations are not perfect and could potentially be improved. The method for clustering and classification could be made more realistic, for example by incorporating other general categorisation principles, such as maximising distances between different colour categories as suggested by Jameson and D'Andrade (1997) and Regier et al. (2007). As the chromatic environment does not fully explain colour typology, especially not the remaining discrepancies between the resulting centroids and human colour categories or the cross-cultural similarities as observed in the World Color Survey (Kay et al., 2003), we would like to propose that culture has an influence on the acquisition of colour categories. Language would seem to be crucial in this process, as language — and specifically colour words — acts as a catalyst to shape one's colour categories. Steels and Belpaeme (2005) present a computational model in which a population of individuals communicating about colour using a simple language all arrive at shared colour categories. Recently, this has been replicated and extended by (Puglisi et al., 2008), where it was shown that a simple negotiating the meaning of perceptual categories, such as colour categories, will result in a self-organised communication system using a limited set of words and a limited set of shared categories. Furthermore, if the agents are given a human-like perceptual model, the colour category typology resembles that observed in the World Color Survey

(Belpaeme & Bleys, 2005). Dowman (2007) comes to a similar conclusion, also using a model were individuals learn colour categories and terms culturally.

These computer model results tie in with anthropological observations on how language and culture impacts on the nature of colour systems (Davidoff et al., 1999; Roberson, 2005) and on psychological results showing the influence of colour words on colour perception (Roberson et al., 2008).

References

Belpaeme, T. & Bleys, J. (2005) Colourful language and colour categories. In A. Cangelosi and C. L. Nehaniv, editors, *Proceedings of the Second International Symposium on the Emergence and Evolution of Linguistic Communication* (EELC '05), Hertfordshire, UK: The Society for the Study of Artificial Intelligence and the Simulation of Behaviour.

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

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

Bradley, P. And Fayyad, U. M. (1998) Refining initial points for k-means clustering. In *Proceedings of the 15th International Conference on Machine Learning*, pages 91-99, San Franciso, CA: Morgan Kaufmann.

Brainard, D. H. (2003) Color appearance and color difference specification. In S. Shevell, editor, *The Science of Color*, pages 191-216. Optical Society of America.

Burton, G. & Moorhead, I. R. (1987) Color and spatial structure in natural scenes. *Applied Optics*, 26, 157-170.

Conover, W. (1999) Practical nonparametric statistics. New York: John Wiley and Sons.

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

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

Fairchild, M. D. (1998) Color Appearance Models. Reading, MA: Addison-Wesley.

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

Howard, C. M. & Burnidge, J. A. (1994) Colors in natural landscapes. *Journal of the*

Society of Information Display, 2, 47-55.

Jameson, K. & D'Andrade, R. (1997) It's not really red, green, yellow, blue: an inquiry into perceptual color space. In C. Hardin and L. Maffi, editors, *Color categories in thought and language*, pages 295-319. Cambridge, MA: Cambridge University Press.

Kay, P. & Regier, T. (2003) Resolving the question of color naming universals. *Proceedings of the National Academy of Sciences of the USA*, 100, 9085-9089.

Kay, P., Berlin, B. & Merrifield, W. (1991) Biocultural implications of systems in color naming. *Journal of Linguistic Anthropology*, 1, 12-25.

Kay, P., Berlin, B., Maffi, L. & Merrifield, W. R. (2003) *The World Color Survey*. Stanford: Center for the Study of Language and Information.

Komarova, N. & Jameson, K. (2008) Population heterogeneity and color stimulus heterogeneity in agent-based color categorization. *Journal of Theoretical Biology*, 253, 680-700.

Lindsey, D. T. & Brown, A. M. (2002). Color naming and the phototoxic effect of sunlight on the eye. *Psychological Science*, 13, 506-512.

Lloyd, S. P. (1982) Least squares quantization in pcm. *IEEE Transactions on Information Theory*, 28.

Neitz, J., Neitz, M. & Jacobs, G. H. (1993) More than three different cone pigments among people with normal colour vision. *Vision Research*, 33, 117-122.

Puglisi, A., Baronchelli, A. & Loreto, V. (2008) Cultural route to the emergence of linguistic categories. *Proceedings of the National Academy of Sciences of the USA*, 105, 7936-7940, 2008.

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

Roberson, D. (2005) Color categories are culturally diverse in cognition as well as in language. *Cross-Cultural Research*, 39, 56-71.

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, 752-762.

Rosch-Heider, E. & Olivier, D. (1972) The structure of the color space in naming and memory for two languages. *Cognitive Psychology*, 3, 337-354.

Sharpe, L. T., Stockman, A,. Jägle, H. & Nathans, J. (1999) Opsin genes, cone photopigments, color vision, and color blindness. In K. R. Gegenfurtner and L. T. Sharpe, editors, *Color Vision: From genes to perception*. New York: Cambridge University Press.

Shepard, R. N. (1992) The perceptual organization of colors: An adaptation to regularities of the terrestrial world? In J. Barkow, L. Cosmides, and J. Tooby, editors, *Adapted Mind*, pages 495-532. Oxford: Oxford University Press.

Steels, L. & Belpaeme, T. (2005) Coordinating perceptually grounded categories through language. A case study for colour. *Behavioral and Brain Sciences*, 24, 469-529.

Sturges, J. & Whitfield, T. A. (1995) Locating basic colours in the Munsell space. *Color Research and Application*, 20, 364-376, 1995.

Van Wijk, H. (1959) A cross-cultural theory of colour and brightness nomenclature. *Bijdragen tot de taal-, land-en volkenkunde*, 115, 113-137.

Winderickx, J., Lindsey, D., Sanocki, E., Teller, D., Motulsky, A. & Deeb, S. (1992) Polymorphism in red photopigment underlies variation in color matching. *Nature*, 356, 431-433.

Wyszecki, G. & Stiles, W. (1982) *Color Science: Concepts and Methods, Quantitative Data and Formulae*. John Wiley and sons, New York, 2nd edition. Reprinted in 2000.

S. N. Yendrikhovskij. (2001a) Computing color categories from statistics of natural images. *Journal of Imaging Science and Technology*, 45, 409-417.

Yendrikhovskij, S. N. (2001b). A computational model of colour categorization. *Color Research & Application*, 26, 235-238.

Appendix I. Conversion from *RGB* to *XYZ*

For the conversion from RGB to *XYZ*, viewing conditions were assumed to be a PAL/SECAM CRT-screen using D65-light as reference white ($X_n = 95.04$, $Y_n = 100.00$, $Z_n = 108.88$) and using $\gamma = 2.5$. The conversion matrix for $XYZ = RGB^{\gamma}$. M is defined as:

$$M = \begin{bmatrix} 0.430587 & 0.222021 & 0.0201837 \\ 0.341545 & 0.706645 & 0.129551 \\ 0.178336 & 0.0713342 & 0.93934 \end{bmatrix}$$

Equations to convert *XYZ* to CIE L u v or L a b can be found in, among others, (Fairchild, 1998).

Appendix II. Computing Hue and Chroma Values

The hue and chroma values can be computed using Equations 1 and 2, respectively in the CIE $L^*u^*v^*$ and CIE $L^*a^*b^*$ colour space.

$$H_{AB} = \arctan \frac{b*}{a*}, \quad C^*_{AB} = \sqrt{a*^2 + b*^2}, \tag{1}$$

$$H_{AB} = \arctan \frac{v*}{u*}, \quad C^*_{UV} = \sqrt{u*^2 + v*^2}, \tag{2}$$