Colour associations in a young adult Indian population

Valérie Bonnardel[1, 2] Nijoo Dubey[1], Sucharita Beniwal[1], Mayukhini Pande[1] and Clare Davies [2]

Abstract. Colour possesses a powerful symbolic function that might predate the emergence of language and its use as a sign is prevalent in almost any human activities. The nature and the underlying structure of word-colour associations are of interest for colour theory, semiotic and aesthetics. In this study, data reduction and visualisation analyses i.e., exploratory factor analysis and Multidimentional scaling were performed with the aim to identify word-colour association patterns. In a free association task, thirty-seven Indian young adults spontaneously associated words to eight colours (spaning over the colour circle) providing a total of 1333 associations.
Three different structures were identified. A first structure extracted from Natural Object word-colour associations revealed three components: [Bluish – Water], [Greenish – Vegetation] and [Yellowish/Red - Fruits] is evocative of archetypical associations. The second structure derived from Concept word-colour associations correspond to [Cool - Positive Emotion Low Energy] vs. [Warm - Religion and Spirituality] colours with an additional third [Blue/Purple – Lifestyle] component and reflects socio-cultural stereotypes. The third structure extracted for Man-Made Objects word-colour associations is undefined in terms of colour order system, the association seem to be idiosyncratic to the sample of participants tested.

1 INTRODUCTION

Archeological records show that ochre was mined and used specifically for its colour and was associated to mundane activities as early as the Middle Paleolithic. These associations suggest ochre to be an early cultural symbolic construct used in non-literate societies [1]. Today, in modern societies, colours convey complex meanings and based on its relationship with the referent, three aspects of its semiotic function can be distinguished: colour acts as (1) an icon when relation of similarity with the referent exists (e.g., blue and sky); (2) as an index when there is a relation of physical continuity (e.g. greyish sky and rain) and (3) as a symbol when the relation is arbitrary, formal or agreed and independent of the referent physical characteristics (green, safety) [2]. Colour as an icon or an index is expected to provide universally shared associations when referents are from natural environment. When colour acts as a symbol its association to a referent results from built cultural consensus for which universality is not a priori expected. There will be other instances where word-colour associations, either as icons, indexes or symbols, will be limited to small group of individuals or even be idiosyncratic to an individual depending on personal experience.

Studies in which words were associated with colours have concluded to both regularities (e.g. red associated with heat) and cross-cultural variations [3]. Often studies have been using a limited and predefined list of words to be associated to colours. This procedure facilitates the production of high frequency word-colour associations from which stereotypes are identified, but it rules-out the possibility to collect spontaneous and idiosyncratic associations. Using a free association task, the aim of the present study was to use data reduction and visualization from an unrestrained corpus of word-colour associations with the objective to identify structures underpinning colour symbolism system.

2 METHOD

2.1 Subjects

Thirty-seven subjects (17 F) mean age 34 years (SD 7.7) Ahmedabad (Gujarat, India) participated to the experiment. Eight subjects were young Indian Professional Population, who were educated at post-graduation level or similar professional Studies and who were working in the Corporate Sector of the Indian Economy and received an emolument of Rs. 600 per person for their participation. Twenty-nine subjects (10F) were recruited by the researchers among their acquaintance; they participated on a voluntary and unpaid basis for their own interest in the experiment. All subject had normal color vision as assessed by Ishihara color test.

2.2 Stimuli

Selected samples from A6 NCS box (NCS 1950 original colors) were used. Four samples, as judged by the experimenters to best represent 'Red' Y90R (SS 1085-Y90R) [R], 'Blue' R90B (SS 3065-R90B) [B], 'Green' G10Y (SS 2070-G10Y) [G] and 'Yellow' (SS 0580-Y) [Y] – Four additional samples were chosen as intermediate between two primary basic colors: 'Orange '(SS 0585-Y50R) [YR] , 'Purple' SS 4050-R50B [RB], 'Turquoise' SS 3060-B40G [BG] and 'Chartreuse' SS 1075-G60Y [GY].

2.3 Procedure

Subjects were asked to write as many words they could associate to a color sample presented for 60 seconds. The subjects were refrained from using any personal associations (e. g. "my childhood sweater"), and asked not to hesitate from writing any unpleasant words. The subject was then presented with 8 different color samples one at a time under the Macbeth X-rite D65 ceiling panel, on a grey table. Colour samples were shown in a random sequence for each subject and never named by the experimenter.

1. National Institute of Design –Ahmedabad (India)
2. University of Winchester – Winchester (UK)
Valerie.bonnardel@winchester.ac.uk

3 RESULTS

3.1 Word and category frequencies

A total of 1333 words were elicited. Words were grouped into Concepts [C] and Objects with a ratio of 3.2:1 (Table 1). The colour Red elicited the greatest number of concept words and Green the greatest number of Object words. A Chi-square test (Pearson Chi-square = 18.6, df=7, p<0.01) indicated that 'Orange' was significantly associated to [C] words (z-score = 2.6, p<0.01), while 'Green' was associated to Object words (z-score = 2.2, p < 0.05). The four primary basic colours [R-G-Y-B] accounted for 62% of Object words and 55% of [C] words. Females provided a larger number of words compared to males, this difference essentially lay in the number of Object words, but was not significant.

Object words were further categorised into Natural Objects [NO] and Man-made Objects [MO] with a ratio of 3.5:1. Blue colour was significantly associated to Man-made objects (Pearson Chi-square = 18.7, df=7, p<0.01, z-score = 2.6, p<0.01). The four primary basic colours account for 61% and 65% for [NO] and [MO] respectively.

Table 1. Total number of Concept and Object words elicited by the eight colours.

	Total	Concept	Object *	NO	MO
	190	129	61	54	7
	167	132	35	24	11
	225	157	68	51	17
	117	80	37	30	7
	160	102	58	36	22
	138	93	45	35	10
	191	112	79	64	15
	145	99	46	41	5

*Object words are further divided into Natural [NO] and Man-made Objects [MO].

Based on semantic similarities, the 1333 words were grouped in 60 categories: 38 Concept [C], 11 Natural Object [NO] and 11 Man-made Object [MO] categories (Table 2).

'Visual Appearance' is the [C] category eliciting the largest number of words (73) and corresponds to 8% of the overall [C] words with a word repetition index (total number of words /number of unique words) equal to 3.2. 'Vegetation' is the largest [NO] category (75) and represents 22% of the total number of words for this category with a word repetition index equal to 4.7. 'Brand' is the largest [MO] category (15) that accounts for 16% of the [MO] words, each word in this category was a unique word. Words related to 'Public Transportation' and 'Technology' were elicited by males only.

Table 2. Category name, category example, total number of words and number of unique words for [C], [NO] and [MO] categories.

Category	Example	Total nb. of words	Nb. Unique words
Concepts			
Visual Appearance	Saturated	73	23
Positive Emotion Low Energy	Peaceful	64	17
Natural Cycle	Monsoon	48	16
Religion & Spirituality	Islamic art	47	33
Happiness	Enjoyable	45	20
Temperature	Fair atmosphere	37	10
Positive Emotion High Energy	Euphoria	36	19
Striking	Eye catching	36	18
High Energy	Intense	31	13
Freshness	Freshness	29	3
Lifestyle	Designer	28	23
Wealth	Prosperity	26	8
Health & Hygiene	Malnourished	25	16
Negative Emotion High Energy	Jealousy	24	13
Life & Youth	Lifeless	23	8
Nature	Horizon	21	8
Political awareness/ Civic	Love for nation	20	15
Negative Emotion Low Energy	Melancholy	19	14
Not Striking	Leftover	19	12
Depth & Mystery	Deep thoughts	17	6
Material Properties	Sharp	17	14
Motion	Flying	17	15
Positive Social Behaviour	Friendly	17	14
Positive Moral Values	Fairness	16	14
Creation	Creative	15	13
Social Event	Song	15	8
Low Energy	Lethargic	14	10
Social Groups	Children	14	10
Artificial	Cheap	13	8
Different	Individualistic	13	8
Geographic Location	Holland	12	11
Move ahead	Achievements	12	10
Sex	Seductive	12	7
Warning	Stop	12	6
Purity	Untouched	11	7
Attractiveness	Radiant	9	7
Progress	Anti pollution	9	9
Taste	Lemony	8	7
Natural objects			
Vegetation	Algae	75	16
Fruits	Papaya	48	19
Water	Water fall	44	12
Fire	Light	41	6
Atmosphere	Thunder	28	6
Organic	Blood	28	7
Flowers	Marigold	17	11
Vegetables	Coriander leaves	16	9
Earth	Island	15	8
Animal	Peacock	14	5
Grains & Spices	Turmeric	9	6
Man-made Objects			
Brands	Dairy milk	15	15
Clothes	Denim	13	6
Nation	Indian flag	12	6
Religion	Sindoor	10	6
Lifestyle & Accessory	Lipstick	9	6

Signs	Smiley	8	5
Sports	Pakistani cricket team	7	7
Public Institutions	Hospital curtains	6	4
Public Transport	Shatabdi express	5	4
Utility	Ink	5	2
Technology	Radium	4	4

Words and the categories most frequently associated to a given colour are reported in Table 3. 'Blood' is the single most frequently associated word and exclusively so to a colour. 'bright' and 'fresh' were associated 37 and 23 times respectively, but their association spread across different colours.

Words associated to the primary basic colours are 'blood' and 'love' [R], 'tree', 'grass' and 'fresh' [G], 'sun' and 'light' [Y] 'sky' and 'water' [B]. Considering category-colour associations, Vegetation-[G] is the strongest association (45 words), followed by Fire-[Y] (25 words), Organic-[R] (24 words) and Water-[B] (17 words). It is interesting to note that [R] is ambivalently associated to positive (PEHE) and negative (NEHE) emotions, although both share the high energy characteristic (e.g. euphoria, jealousy). On the other hand, low energy positive emotions (LEPE) (e.g. peaceful) is most frequently associated with [G], [BG] and [GY]. This opposition between high and low energy emotions, thus coincides with the red-green colour opponency.

Words associated with secondary basic colours are 'water' and 'sea' [BG], 'fresh' and 'grass' [GY], 'orange', 'fresh', 'sun' and 'vibrant' [YR] and 'royal' and 'grapes' [RB]. 'Vegetation' - [GY] is the strongest association followed by 'Religion and spirituality' - [YR], then 'Water'- [BG] and 'Fruit'- [RB].There are significantly more word-colour associations with primary than secondary basic colours (10.6 (5.0) vs. 5.5 (1.8), t=3.0, p=0.007). Categories too have a larger number of words associated to primary colours (19.7 vs. 14.3) but the difference is not significant.

Table 3. Most frequently associated words and categories to colours

	Word associations		Category associations	
	Word	Count	Category	Count
	Blood Love	22 8	**NO**-Org **C**-PEHE **C**-NEHE (**C**-VisA)	22 13 13 (13)
	Tree Grass Fresh	13 6 6	**NO**-Veg **C**-PELE	45 17
	(Bright) Sun Light	(17) 12 7	(**C**-VisA) **NO**-Fire **C**-Hap **C**-NatC	(28) 25 15 15
	Sky Water	12 10	**NO**-Water **NO**-Atm	17 15
	Royal Grapes	5 4	**NO**-Fru **C**-Life Sty **C**-Wealth	11 7 7
	Water Sea	6 6	**NO**-Water **C**-PELE	19 16
	Orange (Bright) Sun Fresh Vibrant	8 (5) 4 4 4	**C**-R&S **NO**-Fruit	20 10
	Fresh Grass	9 5	**NO**-Veg **C**-PELE	25 14

Visual Appearance [C] category (e.g. 'bright', 'dark') elicited the largest amount of words (73). These words refer to the quality of colour itself where color is conceived as an entity on its own. Since the study's aim was to investigate the structure of the word-colour associations in which color is used as a sign, 'Visual Appearance' category and words are in brackets in Table 3 to indicate their exclusion from the present description and these associations were removed from subsequent analyses.

3.2 Principal Component Analysis

A first exploratory Principal Component Analysis was performed with the overall 59 categories (37 [C], 11 [NO] and 11 [MO]). Factors were extracted with eigenvalue greater than 1 and a Varimax rotation was applied. Three factors; PC1 [G-GY], PC2 [Y-YR-R-RB] and PC3 [B -BG] explain 68% of the total variance (Table 4). The three components correspond to two distinct cool-colour clusters; Green (PC1) and Blues (PC3) and a warm-colour interval (PC2). Inspection of the component scores (calculated with Anderson-Rubin method) indicates that the most representative categories for each PC are from [NO] categories and correspond to 'Vegetation' (PC1), 'Fruit' (PC2) and 'Water (PC3). The cool-colour components (PC1 and PC3) are further associated with Positive Emotion Low Energy [C] category. 'Atmosphere' [NO], 'Flower' [NO] and 'Life Style' [C] have the lowest scores on PC1, and 'Organic' [NO], 'Sex' [C] and 'Warm' [C] have the lowest scores on PC3. 'Public Institutions', 'Utilities' and 'Public Transportation' all from [MO] categories correspond to the lowest scores on PC2. It is worth noting, that [R] and [RB] have the lowest factor loading.

Table 4. PCA applied to the overall data set (N=59) – Percentage of variance for each component, and their scores associated to categories for upper and lower percentiles

Factor loading	% of variance	Above 95% percentile	Below 5% percentile
.95 .95	(PC1) 24	NO-Vgtn C-PELE C-Frsh	NO Atm C-LSty NO-Flw
.81 .83 .45 .48	(PC2) 22	NO-Fru C-NatC NO-Fire	MO-Pubin MO-Uti MO-PubTr
.88 .84	(PC3) 22	NO-Water NO-Atm C-PELE	NO-Org C-Sex C-Warm

Table 5: PCA applied to [C], [MO] and [NO] categories. Same as in Table 4.

Concepts (N=37)			
Factor loading	% of variance	Above 95% percentile	Below 5% percentile
.69 .87 .88	PC1 29	PELE	Sex Warm
.76 .88 .54	PC2 21	Religions & Spirituality	Artificial Attractiveness
.83 .68	PC3 19	Lifestyle	Sex Warm

Man-made objects (N=11)			
Factor loading	% of variance	Above 90% percentile	Below 5% percentile
.93 .62 .88	PC1 29	Branding	Clothes
.72 .88	PC2	Nation	Clothes

			20		
.96			PC3 18	Religion	Public Institutions
.93			PC4 15	Lifestyle & accessories	Sport
Natural objects (N=11)					
.97	.96		PC1 27	Vegetation	Organic
.89	.87		PC2 26	Water	Organic
.73	.94	.70	PC3 25	Fruit	Organic

PCA performed with [C] categories extracted three PCs that accounted for 69.4% of the total variance. One component corresponds to a cool-colour dominance (GY is considered as a boundary between warm and cool colours) for which 'PELE' has the highest score and 'Sex' and 'Warm' the lowest. The second component corresponds to warm-colours with 'Religion & Spirituality' scoring the highest while 'Artificial' and 'Attractiveness' categories score at the lowest on this factor. 'Lifestyle' has the highest score and 'Sex' and 'Warm' have the lowest scores on the third PC that combines [B- RB].

Man-made object PCA extracted four PCs accounting for 83% of the total variance. The first two PCs exhibit a colour opponent structure. PC1 corresponds to [Y-YG- B] for which 'Branding' has the highest score and 'Clothes' the lowest. The highest score on PC2 [YR-G] is for 'Nation' which corresponds to Indian flag's Indian saffron and Indian green colours, 'Clothes' has the lowest scores on this factor. The other two components correspond to single colours: [R] with 'Religion' for highest scores and 'Public institution' lowest score and [RP] with 'Lifestyle & Accessory' for highest scores and 'Sport' for lowest scores. [B] was not extracted by the PCA.

PCA for [NO] extracted three components accounting for 78.5% of the total variance. The two first components are identified as green [G- GY] and blue [B-BG], the last component is a composite of warm-colour with a dominance of yellow [Y-YR-RB]. The two first components are associated to 'Vegetation' and 'Water' and the third to 'Fruit'. It should be noted that 'Organic' category in which the word 'blood', exclusively associated to [R], contributes to 61% of the words for this category is the category with lowest scores for the three component. [R] was not extracted by the PCA.

3.3 Multidimensional Scaling

In this last analysis, data were summarised into a frequency matrix. In this case the eight colour names were the column headings within the term-colour table, with one table row for each of the 620 unique words and phrases elicited across all participants. Thus, each column vector was the list of 620 values for how often each word had been linked to that colour.

This matrix was first weighted, using the most common weighting combination employed in computational linguistics techniques such as Latent Semantic Analysis [4]. This involved recalculating each term frequency based on two mathematical transformations. A 'local' logarithmic weighting, 1+log(original frequency), aimed to reduce the impact of very frequent association , e.g. 'blood' – [R], which could otherwise pull it disproportionately far from the other colours. A 'global' weighting based on the information theory concept of 'entropy' (where a term with more entropy is more effective at discriminating among colours) [5] was intended to reduce the impact of terms which appeared across several colours and thus would discriminate poorly between them. To calculate the similarity matrix among the colours, the cosine measure was used, which is the common choice for term or document similarities in computational linguistics and information science[1]. Following the procedure outlined by Louwerse and Zwaan (2009) [6] and Davies (2013) [7] for extracting a 3D space from the cosine similarity matrix, first Euclidean distances were calculated from the cosine values, and then the 8 x 8 distance matrix was subjected to classical (metric) multidimensional scaling (MDS) to derive a 3D space.

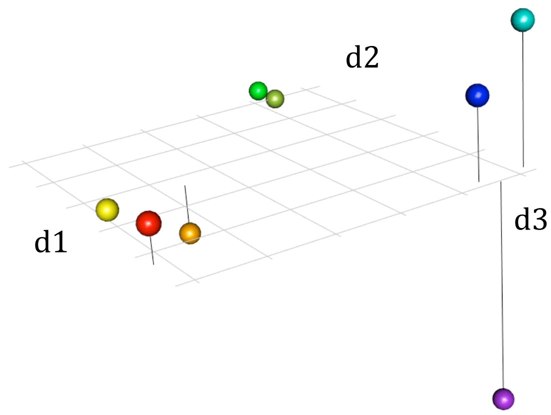

Figure 1: 3D MDS plot

The MDS 3D solution (goodness of fit = 0.71) reveals a three-clusters configuration, showing similarities with the structure extracted from the PCA performed on the 59 categories (Table 4). In 3D, the three MDS clusters correspond to Bluish [B + BG]; Greenish [G+GY] and Yellowish-Red [Y+YR+R] colours, recalling the three primaries of a trichromatic system. [RB] is singled out along the third dimension.

4 DISCUSSION

PCA and MDS performed on the overall word-colour associations suggest a 3D structure; Bluish, Greenish, Yellowish-Red with, as the most representative categories, 'Water', 'Vegetation' and 'Fruits' stemming from Natural Objects categories. It is quite striking that the MDS, which was of course completely unconstrained as to where each colour could emerge within the assumed 3D space, nevertheless separated the colours (in clusters of either one, two or three) into the corners of an approximate cuboid. This shape implies the same near-orthogonal dimensions as the

---

[1] Cosine similarity values are commonly used as a way of comparing the vectors based on the *angles* between them in the original multidimensional space, without being heavily influenced by their actual *lengths* - because the latter depend on raw word frequencies, which here were not of interest in themselves. For instance, with the current data the use of cosines further prevented the unusually frequent, and uniquely red-associated, occurrences of 'blood' from disproportionately distancing red away from all other colours.

PCA, which of course would constrain the data into that pattern.

PCA, performed for each of the three individual categories, reveals a 3D structure for Natural Object and Concepts categories. The [NO] structure is close to that extracted for the overall data set. The [C] structure is better described by a cool-warm colour dimension with [B –RB] as the third component.

The [MO] structure is represented by four components. Two components combine opponent colours [Y-YG-BG] (yellowish vs Bluish) and [YR- G] while two correspond to single colours [R] and [RB]. Category scores that are indicative of their association to each component provide interesting information. For the sample of participants tested, colours associated to ‘Brands’ and ‘Nation’ were not associated to ‘Clothes’; [R] associated to ‘Religious Objects’, was not associated with ‘Public Institutions’ and ‘RB’, associated to ‘Lifestyle Accessories’ was not associated with ‘Sport’.

In conclusion, three structures have been identified. The 3D structure identified for the overall data appears to be driven by the archetypical [NO] word-colour associations in which colour acts as an icon with [Bluish – Water], [Greenish-Vegetation] and [Yellowish/Red-Fruits]. A second 3D structure identified in [C] categories is described as cool [Positive Emotion Low Energy – cool] and [Religion and Spirituality-warm] colour dimension, with a third [Lifestyle – B/RB] component. These associations are established through social and cultural influences and reflect potential cultural stereotypes. The structure emerging from [MO] categories is more complex. It includes both opponent-colour and single colour components with no obvious colour order system structure. The word-colour associations in this category are representative of the sub-culture values that are idiosyncratic to the sample of participant tested, the majority of them were trained in Design which will explain colour association with ‘Branding’ and ‘Lifestyle Accessories’.

REFERENCES

[1] E. Hovers, S. Ilam, O. Bar-Yosef, and B. Vandermeersch, ‘An early case of color symbolism. Ochre use by modern Humans in Qafzeh cave’ *Current Anthropology*, **44**, 491-522. (2003)

[2] J. L. Caivano ‘Color and Semiotics: A two-way street’ *Color Research and Application* **23**, 390-401. (1998).

[3] L.R. Prado-Leon, R. Avila-Chaudran and R. A. Rosales-Cinco, ‘Colour Associations in the Mexican University Population’, *Progress in Colour Studies, Vol II: Psychological aspects*. Amesterdam: John Benjamins Ltd. (2006)

[4] A. Lifchitz, S. Jhean-Larose and G. Denhière, ‘Effect of tuned parameters on a LSA multiple choice questions answering model’, *Behavior Research Methods* **41**, 1201-1209 (2009).

[5] C.E. Shannon, ‘A Mathematical Theory of Communication’, *Bell System Technical Journal* **27**, 379-423 (1948).

[6] M.M. Louwerse and R.A. Zwaan, ‘Language Encodes Geographical Information’, *Cognitive Science* **33**, 51-73 (2009).

[7] C. Davies, ‘Reading Geography Between the Lines: Extracting Local Place Knowledge from Text. In: T. Tenbrink, J. Stell, A. Galton and Z. Wood (eds.)’, *Spatial Information Theory* (Lecture Notes in Computer Science Vol. 8116). Heidelberg: Springer, 320-337 (2013).