Free-sorting of colors across cultures: Are there universal grounds for grouping?

Debi Roberson

University of Essex

Ian R. L. Davies

University of Surrey

Greville G. Corbett

University of Surrey

Marieta Vandervyver

University of Windhoek

Running head: Free-sorting of colors

Keywords: grouping; color categories; free-sorting; universality; cultural relativity

Author's Note: Debi Roberson, Department of Psychology, University of Essex, UK; Ian Davies, Department of Psychology, University of Surrey, UK; Greville Corbett, Department of Linguistics, University of Surrey, UK; Marieta Vandervyver, Department of Nursing, University of Windhoek, Namibia. Correspondence concerning this article should be addressed to: Dr. Debi Roberson, Dept. of Psychology, University of Essex, Wivenhoe Park, Colchester, UK, CO3 4SQ. Tel:01206 873710, Fax: 01206 873590, email: robedd@essex.ac.uk.
The experimental studies reported here were partly supported by ESRC grant No. R000236750 to Davidoff, Davies, and Corbett; ESRC grant No. R000238310 to Davidoff, Roberson and Davies and by a University of Essex RPF award to the first author. Some of the data (English, Russian and Tswana) has been reported in a different form (Davies & Corbett, 1977). We are grateful to Jules Davidoff, Goldsmiths College, University of London, in collaboration with whom some of the research reported here was carried out and to the following who collected data: Syd Hiskey (!Xoo); Tat'jana Borisovna Sosenskaja, Anna Rum_iskaja and El'zara Orud_evna Ibragimova (Tsakhur); Tat'jana Borisovna Sosenskaja, Pavel Grashchenkov, Isa Magomedov and Madina Magomedova (Bagwalal) .

These studies examined naming and free-sorting behavior by informants speaking a wide range of languages, from both industrialized and traditional cultures. Groups of informants, whose basic color vocabularies varied from 5 to 12 basic terms, were given an unconstrained color grouping task to investigate whether there are systematic differences between cultures in grouping behavior that mirror linguistic differences and, if there are not, what underlying principles might explain any universal tendencies. Despite large differences in color vocabulary, there were substantial similarities in grouping behavior across language groups, and substantial within-language variation across informants. It seems that all informants group stimuli based on some criterion of perceptual similarity, but those with large color vocabularies are more likely to group stimuli in line with their basic color terms. The data are best accounted for by a hybrid system that combines a universal principle of grouping by similarity with culture-specific category salience.

A number of authors have noted the human compulsion for grouping things in the world into categories (Malt, 1995; Schyns, Goldstone & Thibaut, 1998, Roberson, Davidoff, Davies & Shapiro, 2004). Indeed categorization seems to be a fundamental part of human cognition. Young children start to systematically and exhaustively sort groups of similar looking objects by spatial location at approximately the same time (within 10 days) as they enter the 'naming spurt' (a sudden sharp increase in vocabulary at around 18 months) (Gopnik & Meltzoff, 1997). This link between exhaustive sorting and naming has also been found to occur for children with Downs syndrome (Mervis & Bertrand, 1994). So it appears that noticing that things in the world can be classified into groups promotes fast word learning (particularly of count nouns that label kinds of objects) and that label learning, in turn, supports the urge to categorize.

Although there are obvious advantages to such grouping behavior for cognitive economy, inference making and interaction with the world (Rosch, 1975), the basis for such groupings is still the source of considerable controversy (Steels & Belpaeme, in press; Levinson, Kita, Haun & Rasch, 2002; Saunders & van Brakel, 1997). It could be that there are some obvious natural groupings in the world that human perceptual systems cannot help but notice, as suggested by Rosch, (1973) in which case human categorization would merely mirror the divisions already present in the world; or that particular cultural needs and knowledge systems drive different groups of individuals to make different groupings, in which case some groupings would be more likely in certain conditions than others (Wierzbicka, 1990). Yet a third possibility is that some combination of natural discontinuities and particular needs and goals operates to produce hybrid systems of categorization, with a universal set of underlying constraints (Malt, 1995).

In seeking to disentangle the roles of knowledge, goals and natural salience in categorization a number of researchers have investigated the domain of color. Whilst the color dimension is a perceptual continuum within which humans can detect millions of just-noticeable-differences of hue, brightness and saturation, (Brown & Lenneberg,

1954), there is considerable diversity in the way that different cultures segment the continuum of visible colors linguistically. Some languages have been reported to use as few as two 'basic' terms to describe all visible colors (Rosch-Heider, 1972). Others have been reported to use between three and eleven (Berlin & Kay, 1969), while some have twelve (Russian; Davies & Corbett, 1997; Paramei, 2005) or more. Kay, Berlin and Merrifield (1991) describe 'basic' terms (BCTs) as those terms that are monolexemic, present in the idiolect of all observers and not subsumed within the meaning of other terms. Once one considers secondary terms there is far greater diversity (English has some 4,000 words or phrases to describe colors (Brown & Lenneberg, 1954). However, within these diverse naming systems there are some noticeable generalities (Kay, Berlin, & Merrifield, 1991; MacLaury, 1997). It is the finding of such generalities that has led to the proposal that color might be one area of experience where natural discontinuities arise (through the properties of the visual system) that lead to universals in cognitive color categorization that transcend terminological differences (e.g., Heider & Olivier, 1972).

A number of recent studies have investigated measures of naming, memory and perceptual similarity judgments across cultures with different numbers of linguistic color categories (Davidoff, Davies & Roberson, 1999; Roberson, Davies & Davidoff, 2000; Roberson, Davidoff, Davies & Shapiro, 2004, 2005; Özgen & Davies, 1998; Jameson & Alvarado, 2003a). These studies have found consistent differences in a range of perceptual and memory tasks, systematically linked to the color name categories in each culture. Most recently, Roberson, Davidoff, Davies & Shapiro (2005) have shown that even though two coding systems may appear superficially very similar, speakers of the two languages encode, remember and discriminate color stimuli in different ways. Himba, a language spoken by a semi-nomadic, cattle herding people in South West Africa, shows similarity in the number of linguistic categories for color to Berinmo, the Papua New Guinean language previously studied by Roberson et al. (2000). Both languages have five basic color categories, according to the criteria of Kay et al. (1991). However, Himba participants showed categorical perception only for their

own linguistic categories and not for either the supposed universal categories of English or to those of the Berinmo language.

One criticism of these studies, however, is the suggested possibility that participants routinely recruit language to the tasks used (particularly those involving memory) and thus they do not tap nonlinguistic representations of color (Munnich & Landau, 2003). This is particularly problematic for judgments of similarity amongst patches of color equated for variance in hue, lightness and saturation, when having a similar name might be considered just what makes two items most similar. In addition to a range of tasks requiring strictly prescribed judgments we have, for some time, collected naming and free-sorting data for a range of 65 Color-aid stimuli spaced across the range of visible colors and varying in hue, lightness and saturation. In the free-sorting task, individuals are asked merely to group the items in whatever manner they see fit.

Since the range of stimuli is large and includes both good and marginal examples of each cultures set of categories, one possible way of carrying out the task would be to group together those tiles that would be given the same name. This type of behavior should lead to different numbers of groups, depending on the number of categories the culture uses. If groupings were made only according to basic terms one would expect broad similarity of grouping by individuals within a community, but systematic variation between communities with different numbers of basic categories. However, if grouping is based on some other criterion (e.g. some level of perceptual similarity) then one might expect more individual variation within populations, but less variability between different language groups. We here compare naming and free-sorting behavior for a wide range of languages, from both industrialized and traditional cultures to investigate whether there are systematic differences between cultures in grouping behavior and, if there are not, what underlying principles might explain any universal tendencies. We included two separate samples of (UK) English speakers, the first from the general population and the second from the student population and two samples of Nama-Damara (Khoisan from Namibia) to provide some indication of within-language variation. We knew from previous studies, (Berinmo: Roberson et al.; 2002; Damara: Davies, Roling,

Corbett, Xoagub & Xoagub; 1998; Himba: Roberson et al. 2004; Russian: Davies & Corbett; 1994; Tsakhur: Davies, Sosenskaja & Corbett; Tswana: Davies, MacDermid, Corbett, McGurk, Jerrett, Jerrett, & Sowden; 1992; Turkish: Özgen, Davies; 1998;) pilot work, or from our language consultants, that the number of basic colour terms varied from five to twelve, and these differences were most marked in the blue-green, and red-yellow regions. Aggregated groupings from informants in each language group were compared using Multi-dimensional scaling. Aggregating across individuals might mask individual differences within the groups (and, in particular the possibility that some individuals might adopt strategies of either 'lumping' or 'splitting the stimuli (see e.g., Shaver et al., 1987; Alvarado, 1998) and we examine this possibility. However, the advantage of aggregation is that the data matrix is strongly metric and allows a variety of robust analyses. We return to this issue in the discussion

THE LANGUAGES

Samples were taken from native speakers of the following languages: Bagwalal and Tsakhur (both Caucasus), Berinmo (Papua New Guinea), English (UK), Russian, Turkish; eleven southern African languages: Shona (Zimbabwe) Tswana (Botswana) and Damara, Herero, Himba, Kwanyama, Nama, Ndonga, Kwangali, Mbukushu and !Xoo all from Namibia.

The Bagwalal language belongs to the Andi subgroup of Avar-Ando-Tsez group of the Nakh-Daghestan languages. There are some seven thousand native speakers and there is no written version of the language. Tsakhur is a member of the Lezgic group of the Nakh-Daghestanian family and there are around 30,000 native speakers Ibragimov (1990:3; see also Kibrik, 1999; and Davies, Sosenskaja & Corbett, 1999).

Berinmo (also described as Bitara, e.g., Dye, Townsend and Townsend, 1966) is one of the Alamblak languages of the Sepik Hill family of Papua New Guinea. The language is spoken in three villages, Bitara, Kagiru and Sio, situated on the April and Wogamush rivers. Population of the three villages is estimated to be around 500 individuals.

Shona and Tswana are both Bantu spoken in Zimbabwe and Botswana (central zone S). Damara and Nama are southern Khoisan languages from Namibia (see Davies, Roling, Corbett, Xoagub. & Xoagub, 1998 for a description of their color terms). It is commonly referred to as Nama-Damara, reflecting its common use by two different ethnic groups, the Nama and the Damara (see Malan, 1995). We treat them separately as a further measure of within-language variation. !Xoo is also a southern Khoisan language. The other languages are all Bantu from Namibia. Ndonga and Kwanyama (both Central, Zone R) are the two main languages of the Owambo from the north west. Kwanyama is also spoken in Angola. Herero is the language of the Herero people. Herero and Himba are also Central, Zone R, but in a different branch to the Owambo languages. Mbukushu and Kwangali are both spoken in Kavango, in the north east, and the majority of speakers live in Angola or Zambia. Both languages are Central, Zone K.

METHOD

Participants (by language)

The sample sizes, the composition of the sample by sex and the mean ages (in years) are shown in table 1: English group1 were all students from the University of Surrey, UK. English group 2 were volunteers from a non-student population of normal adults living in Surrey, UK. All Russian participants were volunteers from a population of normal adults living in Moscow, Russia. Older Bagwalal informants had learned some English at school, whereas a few of the younger ones had learned Arabic. All Tsakhur participants were first language Tsakhur speakers, but they also spoke Russian. All Turkish participants were first language Turkish speakers. All Berinmo participants were monolingual Berinmo speakers. The Namibian informants were from rural areas, and most had little or no formal education. They were all native speakers of the language in question, but in some samples, many also knew some Afrikaans. This is reflected in some samples' use of loan terms. All participants had normal color vision (City University Color Vision Test, Fletcher, 1988).

(Table 1 about here)

Interviewers

Russian speakers were tested by either a native speaker or a fluent Russian speaker. Bagwalal and Tsakhur speakers were tested by native speakers, as were the English and Turkish samples. Berinmo speakers were tested by an English speaking experimenter (first author), with the aid of an interpreter from Tok Pisin (New Guinea pidgin) into Berinmo, using back-translation. For the African languages (except Himba and !Xoo), informants were tested by first-language speakers of the appropriate language. For Shona and Tswana, interviewers were trained experimenters. For the remaining African languages, interviewers were student-nurses, studying in Windhoek. They received instruction in data collection from Vandervyver (their tutor) and Davies. Data collection took place in students' native villages, when they returned to their home regions to practice nursing, as part of their normal training. Post-test debriefing was conducted by Bester and Davies, to review data collection procedures. Variation in sample sizes for each language reflects the different number of speakers of each language. Thus four Ndonga speaking interviewers each tested 20 informants, while, there was just one interviewer for each of Kwanyama, Kwangali, and Mbukushu. Himba speakers were tested by an English speaking experimenter (first author), through a Himba-speaking interpreter. !Xoo speakers were also tested by an English speaking experimenter through a !Xoo-speaking interpreter.

Stimuli and Apparatus

The stimuli used were sixty-five Color Aid matt surface colored squares, measuring 2 inches square and backed with stiff card. Best examples of the eleven basic color terms of English (*black, white, gray, red, yellow, blue, green, orange, pink, purple* and *brown*) were included in the set. The stimuli were chosen to sample evenly across the three dimensions of color space (hue, lightness and saturation). Color Aid stimuli were chosen both because they sample across the full range of saturation (the usual Munsell-chip array used to elicit color terms cross-culturally contains only maximally saturated stimuli) and for practical considerations, given the large number of sets required, but

Appendix A gives the Color Aid designations and CIE Y x y and L*a*b* co-ordinates for each color , so that they can be equated to Munsell samples.

Procedure

Participants carried out the tasks in natural daylight, sitting at a table either out of doors, in shaded natural sunlight, or indoors close to a window. Free-sorting always preceded naming to avoid introducing a name-grouping bias. The stimuli were spread out on the table, in random order and the participant was asked to group them so that ones that looked similar were placed together in the way that members of a family go together. Participants who asked for clarification were told that there was no right or wrong way to complete the task, that they should just put together the tiles that they felt should be grouped together. After participants completed the sorting task the groupings were recorded by the experimenter. Subsequently stimuli were presented to participants, one at a time, in random order and participants were asked to name each stimulus.

RESULTS

Color terms in the languages sampled

Table 2 shows the number of Basic color terms (BCTs) for each language by the criterion of Kay et al. (1991). BCTs are monolexemic, not subsumed under the meaning of other terms, not restricted to a narrow class of objects and understood by all observers. English has eleven BCTs: *black, white, gray, red, blue, green, yellow, pink, orange, purple and brown*. Russian and Turkish both have twelve; eleven of them are similar to English, but they divide the blue region into two (dark and light blue; see Moss, Davies, Corbett & Laws, 1990; Paramei, 2005 and Özgen & Davies, 1998 for reports). Tsakhur has a BCT equivalent to the secondary term *turquoise* in English (Davies et al., 1999). Tsakhur and Bagwalal each have a single term for the purple/pink region. Tsakhur is focused in pink and Bagwalal in purple. In addition Bagwalal has no BCTs for orange, brown or gray. Each of these regions is named with another BCT (e.g., orange is named as red).

Some of the African samples had a full set of eleven basic color terms that included various loan terms from colonial languages (Afrikaans, English or German). Thus, Damara, Herero, Nama and Ndonga had eleven BCTs with borrowed terms for pink orange and purple, such as *otjiblou* 'blue', *otjigroen* 'green', *otjipinge* 'pink' and *otjiperse* 'purple' for Herero. Unusually, (see for instance, Davies on Tswana; Davies & Corbett on Ndebele; Davies & Corbett on Xhosa) all the African languages, except Shona, Tswana, Himba and !Xoo, had separate terms for blue and green, which in some cases were loan terms, as in Herero above, but in others they were original terms, such as p*gama* 'blue' *!am* 'green'. Kwangali, Mbukushu, Tswana, Shona, !Xoo, Himba and Berinmo all lack BCTs for pink, purple, and orange. All extend their blue, green/ (or grue) terms to colors that would be called *purple* in English and their red terms to colors that would be called *pink* or *orange*. Berinmo and Himba color naming have been reported in detail elsewhere (Roberson, Davies & Davidoff, 2000; Roberson et al., 2004) as has Damara (Davies, 1998), Tswana (Davies, MacDermid, Corbett, McGurk, Jerrett, Jerrett, & Sowden, 1992) and Turkish (Özgen & Davies,1998).

We report here only the BCTs for each language, as all informants predominantly used these to describe the stimuli and used them with the greatest consensus and consistency. Use of secondary terms and modifiers was limited (for example, less than 10% of all names for both Berinmo and Himba speakers). There were some observable cultural differences in naming behavior. In particular, African informants left more stimuli un-named than speakers of other languages. Overall, those informants whose language contained the largest number of basic terms (but also from the most technologically advanced cultures) also used the greatest number of secondary terms and modifiers, but this still did not account for more than 20% of total naming.

(Table 2 about here)

Number of groups

Table 2 also shows the mean number of groups formed across respondents for each language, the 95% confidence limits and an estimate of each language's number of basic terms. The most notable features are that Bagwalal speakers (33.6) clearly form

more groups than anyone else, followed by Himba (21.4) and Berinmo (20.3). Tswana (17.3), Nama (15.7) and !Xoo (15.4) come next while the remaining language-samples have means ranging from 10.1 for Mbukushu to 13.7 for English group 1. There is no strong relationship between the number of groups and the number of BCTs, although there is a non-significant trend for the languages with the lowest number of BCTs to form the most groups (r = -.35, p = .15 two-tailed). Within language groups there is also some variability in the number of groups formed. For instance, Himba participants made between 6 and 35 groups. Closer examination of individual differences within groups revealed that very few individuals in traditional cultures 'lumped' rather than 'splitting' categories. Only one Berinmo informant and 3 Himba informants made less than 15 groups.[1]

Distance matrices

For each language sample, a dissimilarity or 'distance' matrix was constructed, derived from the grouping task. We assumed that the more similar a pair of tiles were, the more likely it was that they would be grouped together. For each pair of tiles, the proportion of the sample that grouped them together was calculated to give a similarity measure. The similarity measure was then inverted to produce a distance measure and these proportional scores were transformed to arcsine of the proportion. Thus, if two tiles were never grouped together the score would be 1.57 (arcsine (1) in radians) and if they were always grouped together the score would be zero. A matrix based on CIE perceptual distance was also constructed where the entries were the Euclidean distance between the points representing each pair of tiles in CIE L* a* b* co-ordinates. This

[1] To control for the possibility that the few individuals who 'lumped' stimuli had a disproportionate influence on the group plots, Berinmo and Himba matrices were also compared after these individuals' groupings had been removed. The increase in Stress in both cases was extremely small (.001 and .003 respectively). The reduction in variance explained was correspondingly small. Thus it does not appear that these individuals unduly influenced the group solution.

space is designed to represent colors along opponent axes such that the L* axis represents the dimension light to dark, a* is the red-green axis and b is the blue-yellow axis. So, for instance, red is positive high a* and green is negative low a*. Yellow is positive high b* and blue is negative low b*. Subsequent testing showed that using the logarithm of CIE distance (logLab) improved correlations, and we use log distance here.

Correlations among similarity matrices

Correlations across language samples for the grouping matrices were generally large and always positive, ranging from $r = .42$ to .93 with a mean of .69, maximum $p < .001$. (Note that while the magnitude of r is informative, statistical significance is much less so. With 2080 entries in each matrix, correlations as low as 0.1 would be highly significant). All of the grouping matrices were also correlated with logLab ($r = .45$ to .75; mean = .59). However, all the correlations among grouping matrices remain positive and moderately large with perceptual distance controlled for ($r = .14$ to .90; mean = .50). Principal component analysis on the 18 grouping matrices found a single common factor that accounted for 70.00% of the variance. All languages loaded heavily on this single factor with the component matrix weights ranging from .60 for Himba to .93 for Damara, Nama and Ndonga. While there is, again, considerable intra-language variability (even between the English informants tested), it does appear that all informants group stimuli according to some common principle. We return to this issue in the discussion.

Multi dimensional scaling of grouping

Our main analysis consisted of fitting the 18 distance matrices to the INDSCAL multi-dimensional scaling model (Kruskal & Wish, 1981; Norusis, 1994). As in MDS in general, the analysis represents the stimuli (in our case the 65 colors) in an n-dimensional space, such that the Euclidean distance among points represents their dissimilarity: the further apart two stimuli are, the less similar they are. INDSCAL tries to find a common space for all matrices, but incorporates differences among the matrices

(languages) in terms of the relative importance (weights) of each dimension. Thus, each dimension can be 'squashed' or 'stretched' to accommodate differences among the languages and the relative importance of each dimension in the overall solution is given. If the dimensions are interpretable in terms of some familiar color space, then the relative importance of the color space dimensions for each language can be assessed. INDSCAL allows the 'seeding' of the analysis with an initial color space, and here we use CIELab. If CIELab were as good a fit as INDSCAL could find to the original data, then the resultant dimensions would be identical to the seed. On the other hand, if a better fit could be found by re-scaling the original, the resultant dimensions would differ from CIELab to some extent. The number of dimensions is a free parameter in INDSCAL. The higher the number of dimensions, the better the fit to the original data, as indicated by R^2 and Kruskal's stress. However, goodness of fit needs to be tempered by interpretability, and by diminishing returns as the dimensionality increases.

We first applied INDSCAL to the 18 distance matrices for the full set of 65 stimuli. We then 'zoomed in' on three sub-regions where the differing patterns of naming across languages suggested that if there were to be grouping differences related to the language differences, then these were the most likely places to detect them. These three sub-regions were: purple-blue-green; pink-purple; and red-orange-pink. The stimuli for the sub-analyses were selected on the basis of their CIELab co-ordinates, and the predominant name for all languages was either one of the terms used in that region, or they were not named. For instance, for purple-blue-green, tiles were named with either a purple, blue, green or grue term by at least 20% of each sample. CIELab was used as the starting configuration for all analyses and the 3d solution had acceptably small stress levels in all cases, plus the benefit of interpretability of the dimensions. The analyses were also done with no seed, but in all cases the CIELab seed led to lower stress levels.

For each analysis, we mapped the locations of the stimuli in CIELab space (a*, b* and a*, L*); then in the derived dimensions of best fit (dimension1 versus dimension2; dimension1 versus dimension3); and finally, plotted each language in 'weight space' showing the relative importance of the three dimensions for each language. Where there

is clear correspondence between a CIELab dimension and a derived dimension, where possible, we used equivalent axis orientation and we label the axes with their nearest CIELab equivalent. [2]Among our stimuli we labeled the best examples of the English terms as 'landmarks', although there is some variation from graph to graph because of overlap in locations in some views. We also add some tile labels in some graphs to aid interpretation further.

INDSCAL for all 65 tiles

Figures 1a and 1b the 65 Color Aid tiles used in the free sorting task plotted in the 3 dimensions of CIELab a*(red-green), b* (blue-yellow) and figures 1c and 1d shows them plotted in dimension 1 (dim1) and dimension 2 (dim2) of the INDSCAL solution plotted in a*, L* (lightness). The nearest equivalent in Lab for each dimension is given in parenthesis. The achromatic stimuli, black gray and white occupy more or less the same location in a* b* and are labeled gray; brown is not shown but also falls in about the same location. The separate location for these terms can be seen in the a* (red-green), L*(lightness) plane (figure 1b). Figures 1e and 1f show the relative weights for each language for the INDSCAL solution corresponding to the derived dimensions of best fit (d1wt, d2wt). The nearest equivalent in LAB for each dimension is given in parenthesis. The points should be thought of as the ends of vectors, such that the vector length represents goodness of fit for that language to the derived stimulus space, and the angle of the vector represents the relative importance of the two dimensions.

The 3d solution had moderate stress levels for each language (.21-.31). Although the initial configuration had been modified somewhat, the derived dimensions were still highly correlated with CIELab (minimum r = .82). Comparing figures 1a - 1e it can be seen that the stimuli in the derived dimensions (1b and 1c) are more noticeably clustered than in CIELab (1a and 1b) and these clusters tend to include the good examples of the putative universal categories labeled blue, green etc.. The achromatic stimuli (black, gray

[2] In all graphs we exclude the origin to magnify the region of interest, but the continuation of the diagonal from the origin can be constructed by joining the false origin (bottom left) to the top right.

and white) that were not separated in a* b* are more separated in the first two derived dimensions with white and gray occupying the centre, but black being placed close to brown, near yellow. As a corollary, to compare two languages, the angle between their vectors is an index of similarity: the smaller the angle, the greater the similarity. For instance, goodness of fit is low for Himba, and higher for Berinmo, but the relative importance of the dimensions is approximately the same for the two groups. In both cases they weight the red-green dimension more heavily than the blue-yellow dimension, as indicated by their location below the diagonal (equal weights) and the relatively small angle between the two vectors. Kwanyama is similar to Berinmo and Himba, with the remaining languages having relatively small angular separations. Figure 1e shows that most languages weight the red-green dimension more than the blue-yellow dimension as most points lie below the diagonal. The Himba, Russian, Turkish Tswana and the two English groups appear to show this pattern most extremely.

(Figures 1a, b, c, d, e, f about here)

The purple-blue-green region

There were 21 stimuli within a sector below a diagonal joining a*= 50 to b*= 60 and these can be seen in Figures 2a, b with landmark PURPLE, BLUE and GREEN labels. This region is of special interest since five of the languages tested: Tswana, Shona, Himba, !Xoo and Berinmo name this region of the color space with a single term. All other languages have separate terms for green and blue. In addition, Turkish and Russian have two basic blue terms and Tsakhur seems to have a turquoise, hence these languages differentiate the blue green region more than others. Finally, Kwangali appears to have two green terms.

The stresses for the 3d solution ranged from .12 for Damara and Ndonga to .27 for Himba and Shona (mean = .20). The first dimension (weighting = .45) correlated strongly with b*(blue-yellow) (r = .85) while the second most important dimension (.27) correlated strongly with a*(red-green) (r = .85). The third dimension was relatively unimportant on average (.04) and correlated most strongly with L*(lightness) (r = .60). Figure 2c shows the location of the 21 stimuli in the first two derived

dimensions. As with the first analysis, the stimuli are more notably clustered in the derived dimensions than in CIELab. There are three relatively isolated clusters, one in the green region (top left) one in the blue region (bottom left) and one in the purple region (on the right). There are also clusters around green, blue and purple in the other plane (Figure 2d). Figure 2e shows the corresponding language weights for the first two dimensions. There is considerable variation in both the goodness of fit (vector length) and in the relative importance of the two dimensions. The fit is relatively poor for Himba, Berinmo and Shona, and relatively strong for Kwanyama, Ndonga, Nama, Mbukushu, Herero, Damara, English1 and Kwangali. The languages with the highest relative weights for dimension1(~blue-yellow) are Tsakhur, Turkish, Bagwalal, English, Russian and Kwangali. Note, that, these include all the languages with putative extra BCTs in the blue-green region. At the other extreme are: Himba, Kwanyama, Ndonga, and !Xoo. In the other weight plane (Figure 2f), English, Russian and Tsakhur, weight dimension1(~blue-yellow) much more than dimension3 (~lightness), with Kwanyama and Himba at the other extreme, followed by !Xoo, Bagwalal, Berinmo and Ndonga. Thus there is a correlation between the relative goodness of fit for grouping matrices and the number of BCTs in each language. There are more coherent grouping arrangements by informants from those languages with most terms for colors in this region and least agreement from informants whose languages use a single term to denote all colors in this range.

(Figures 2 a, b, c, d, e, f about here)

Purple-pink region

There were 14 stimuli in the purple pink region with positive a* values, and b* values less than 50. Their locations in CIELab are shown in Figures 3a,b. This region is of interest because Tsakhur and Bagwalal each have a single term for purple-pink. Tsakhur is focused in pink and Bagwalal in purple. Kwangali, Mbukushu**,** Shona, !Xoo, Himba and Berinmo have no purple or pink terms, and the area is named partly with the red term and partly with blue or grue terms. Stresses for the 3d solution ranged from .14 for Mbukushu to .30 for Himba (mean = .21). However, unlike the earlier analyses, the

derived dimensions each load on more than one CIELab dimension, and the derived dimensions have, on average, about equal weights, between .22 and .23. Dimension 1 correlates strongly with L*(lightness) (r = .92, but also strongly with b* (blue-yellow) (r = .73). Dimension 2 correlates mainly with a* (red-green) (r = .81) but also correlates with b*(blue-yellow) (r = .54). Dimension 3 correlates with a* (r = .68) and b* (r = .56). Dimension 3 appears to be 'chroma' or colorfulness dimension, which in the CIELab space is the root-mean square of a* and b* and is designated c*. Figures 3c, d show stimulus locations in the derived color space. Light pinks are grouped together towards the top right, blue-purples (e.g., BVBHue) are towards the bottom left, away from red-pinks (e.g.. ROSE) towards bottom right, leaving purples at centre bottom. Similar clusters can be seen in the other plane. Figure 3e shows the weights for the first two dimensions (~L*, ~a*) for each language. It can be seen that, with the exception of Himba, there is not much spread in goodness of fit (vector length), but there is in the relative weights (angles) with Tsakhur weighting dimension 2 the most and Kwanyama, the least. Most of the African languages together with Bagwalal and Berinmo, are located above the diagonal from the origin (highly lightness-based grouping), while languages that have separate terms for red, pink and/or purple are located below the diagonal. But, there are also notable inconsistencies: one English group lies above the diagonal and one below; and Shona clearly falls below, and is apart from other African languages. Fig 3f shows the weights for dimension 3 (~c*) versus dimension 1 (~L*). There is less angular spread in this plane than in Figure 3e, and similarities among related languages are also less clear.

(Figures 3 a, b, c, d, e, f about here)

Red-orange-pink

The 12 stimuli were from the top right quadrant of the a*, b* plane (a* >30, b* >0) and their CIELab locations can be seen in Figures 4a,b. Stress levels for the 3d solution ranged from .14 for English1 to .25 for Himba (mean = .20). The first derived dimension correlated strongly with b* (r = .94), the second with a* (r = .87) but also correlated negatively with L* (r = -.76). The third dimension was harder to interpret as

it did not correlate significantly with any CIELab dimension. Some clue may be gleaned from considering the two highest correlations which are with L* (r = .50) and a* (r = -.43); 'light and not-red'. The relative importance of the dimensions was .33, .23 and .16 for the first to third dimensions respectively. Figure 4c shows the location of the stimuli in the first two dimensions. It can be seen that good reds lie to the right, light pinks to the left, and orange lies at the top. In Figure 4d, the other plane is shown, and the two extremes of dimension 3 are dark-red-pink (ROSE) and light pink (pink). Fig 4e shows the weights for each language for the first two dimensions (~b*, ~a*)and Fig. 4f for the first and third dimensions (~b*, ~unidentified). There is considerable spread of the vector angles in both diagrams. In 4e, English1, English2, Herero, Russian, and Tsakhur clearly weight dimension 1 more than dimension2, and Turkish has the next highest ratio. At the other extreme, Kwanyama, Berinmo, Mbukushu, Kwangali and Ndonga clearly weight dimension 2 more heavily than dimension 1. Bagwalal, Shona and Himba also fall below the diagonal. Thus the languages with separate terms for red, orange and pink weight dimension 1 more than dimension 2 and most of the languages with composite red or yellow terms weight dimension 2 more heavily than dimension 1. A similar separation can be seen in Fig. 4f, except that Damara and Nama now cluster with English, Russian, Tsakhur, Herero and Turkish, all having high dimension1 to dimension3 ratios. Kwanyama lies at the other extreme, with Berinmo, Himba, Ndonga, Mbukushu and Kwangali lying on or below the diagonal.

(Figures 4 a, b, c, d, e, f about here)

DISCUSSION

These studies set out to compare the naming of a set of color stimuli with the unconstrained grouping of those same stimuli by individuals from different cultures, whose color vocabulary differs in both the number of BCTs and the range of colors that those terms denote. Previous studies, using more constrained methods (e.g. 2-alternative forced-choice memory tests, same-different judgments, odd-one-out judgments) have found consistent differences between cultures whose languages code the range of visible colors in different ways (Roberson et al., 2000; Roberson et al., 2004, 2005; Pilling &

Davies, 2004). Those studies, however, used narrow sets of very similar stimuli, and naming may have routinely been recruited to perform the tasks, since other variables were strictly controlled. In the current studies, subjects were asked to group a very disparate set of stimuli in any way they saw fit.

The unconstrained nature of the task resulted in some substantial differences in behavior between individuals. The differences between the mean grouping behavior of the two groups of English informants places them further from each other in figures 1e and 1f than either is to Russian or Damara, for instance. Aggregating group data might mask individual differences in grouping strategy within a language group, such as the tendency to either 'lump' large numbers of stimuli together or 'split' them into many very small clusters, but there are several reasons why this is unlikely to account for the group differences found here.

Firstly, if the tendency to adopt either one or the other of these two strategies were randomly distributed across all groups, such individual differences would weaken the differences found between cultures. Only if individual differences vary systematically with language groups could they give rise to the cultural differences noted above. Secondly, the differences across language groups aren't of a general nature, but are to be found specifically where the languages differ most. Thirdly, excluding the languages with large numbers of groups (the 'splitters'), there are language related differences among samples with very similar mean numbers of groups (and standard errors) e.g. Kwanyama-Kwangali-Mbukushu/Tsakhur-English-Russian.

Moreover, if participants generally chose to group stimuli on some broad universal principles, then broad similarity of grouping should be seen across cultures, in spite of the variability in individual behavior. At the same time, if the perceived similarity of stimuli is genuinely influenced by a learned set of categories, then consistent differences should emerge between speakers of different languages, despite the variability in individual behavior.

The results show evidence of both broad generalities of grouping behavior, as indicated by the strength of the MDS fits between languages, and of some systematic

differences. The degree of fit found between cultural groups could arise if all informants operate a loose general principle of grouping by perceptual similarity (the MDS fits correlate highly with the proximity of the stimuli in CIE Lab space). Such a principle need not be strongly categorical, but might preclude the formation of an arbitrary category that includes, say, red and yellow, but excludes orange (Davies, 1998; Dedrick, 1996; Jameson & Alvarado; 2003b; Roberson et al., 1999). Such a constraint can be equated to slicing an apple. This produces a principled division in which, wherever the cuts are made, the likelihood of two adjacent parts appearing in the same slice is high, while the likelihood of two parts from opposite sides of the apple appearing in the same slice diminishes with the number of cuts made. Thus, the potential for variability when few groups are made is much higher than when many groups are made, and not all potentially possible groupings are logically coherent. If participants take both lightness and hue into account, (dividing the apple along two planes), there are considerable constraints on possible groupings. The principle of grouping by perceptual similarity, rather than by name alone, could yield this degree of inter-language agreement between informants with radically different color vocabularies, provided perceptual color space was shared by all samples.

There is also evidence, from Figure 1f, that all languages appear to weight hue more than brightness when grouping stimuli in an unconstrained way, in spite of the fact that those languages with few BCTs have more lightness-based linguistic categories and the relationships between languages remain quite strong when L*a*b* distances are partialled out. One reason for this might be that the use of only three dimensions in the INDSCAL solution might increase apparent similarities to a certain extent (concealing some regional differences that would be apparent in a solution with more dimensions, but there is a body of evidence (see Jameson, 1997) to show that three dimensions represent psychological color space very well.

In spite of the potential limitations of a 3 dimensional approach, alongside the broad similarities observed between languages, detailed examination of particular areas of the stimulus set reveals some systematic variation between informants whose languages

name each area with different numbers of BCTs,. Examination of the grouping of stimuli in the green – blue – purple range revealed that informants from those languages that do not have separate terms for these stimuli show least agreement in their groupings, while informants from those languages with all three terms show the most similar grouping tendencies. Those languages with separate terms for blue and green also tend to divide the blue and green stimuli into separate groups, as evidenced by their high relative weights on dimension 1 (~b* blue-yellow), while those that use a single grue term do not. Similar systematic differences are seen in the red - orange – pink region, where most of the languages that have separate terms for red, orange and pink weight dimension 1 (lightness) more and dimension 2 less than those languages that have no separate terms for this region. Along all three dimensions, in fact, there is clear separation. Those languages with separate terms make very similar groups of stimuli, while those with composite red or yellow terms make much more diverse groupings. A similar pattern emerges from the grouping of stimuli in the purple – pink region, although here the dimensions of best fit are less clearly associated with hue or lightness dimensions, and there is less agreement between languages on the weighting of the relative dimensions. In particular, the means for two groups of English informants are seen to differ in the relative weighting of dimensions 1 and 2. Thus there is less systematic variation in this region between the sorting behavior of groups who name the stimuli with different numbers of BCTs.

The data make important theoretical contributions to the debate on linguistic and cultural relativity on two counts. Firstly, if individuals always grouped stimuli purely according to their linguistic categories, there should be a high level of within-language agreement on groupings, combined with a systematic variation between languages in the number of groups formed. This pattern was not observed. Instead, it appears that informants generally make more groups than they have BCTs, basing their grouping behavior on a looser universal principle of perceptual similarity. The influence of linguistic category emerges in a more subtle way, in the inverse relationship between the number of linguistic categories an informant has and the number of groups they

chose to make. Informants from languages with a large number of BCTs also have a wide vocabulary of secondary terms, for which there is little inter-informant agreement on the referents. These informants may thus still base their groups on linguistic categories, but use a combination of basic and secondary categories in variable combinations. Those informants with very limited color vocabularies may instead abandon their linguistic categories in this task and group only based on a very restricted criterion of perceptual similarity, thus producing many groups with only two or three stimuli in.

Inevitably, the combination of such a broad data set with an unconstrained task creates a considerable degree of noise in the data. The use of aggregated group data and the compression of solutions to just three dimensions might also inflate the apparent similarity in the behavior of individuals from different cultures and language groups. In spite of this, important differences between languages emerge when detailed examination of sub-sets of the stimuli is carried out. Thus in the green – blue – purple and red – orange – pink ranges, the findings of these studies support a consistent linguistic influence on categorization provided by more constrained paradigms. Even under free-sorting conditions there is still evidence that informants group stimuli in line with their language categories. Thus informants seem to combine a universal underlying grouping constraint with differential cultural goals when freely categorizing colors into groups.

Appendix A

Designations of the 65 Color Aid tiles in the C.I.E. Y x y and L*a*b* metrics. Stimuli were measured under D65 (6500 dg K) and viewed under daylight that varied from 5500 – 7500 dg K. Whilst naturalistic viewing conditions vary slightly over time, this is likely to have added 'noise' to the data rather than any systematic confound.

Tile	Y	x	y	L*	a*	b*
Y-HUE	75.776	.475	.448	89.755	12.065	86.433
Y-S2	16.273	.429	.394	47.331	9.719	30.080
YOY-HUE	65.528	.515	.430	84.755	29.975	90.916
YOY-T4	91.615	.392	.375	96.663	10.502	37.444
YOY-S2	42.236	.408	.390	71.033	8.213	36.056
YO-HUE	58.075	.535	.403	80.780	44.362	82.280
YO-T3	78.261	.437	.384	90.899	23.486	49.119
YO-S3	10.559	.406	.364	38.828	10.388	17.793
OYO-HUE	42.236	.569	.367	71.033	61.929	70.749
O-HUE	39.441	.572	.355	69.070	66.050	64.774
O-S1	27.019	.516	.372	58.992	39.644	47.317
O-S3	9.006	.386	.349	35.996	9.195	12.278
ORO-HUE	25.155	.582	.336	57.226	65.945	51.615
ORO-S3	63.665	.472	.371	83.791	39.039	49.856
ORO-S3	52.484	.403	.352	77.570	21.408	26.076
RO-HUE	22.484	.588	.327	54.537	68.135	48.192
RO-T3	55.280	.483	.346	79.202	51.306	41.374
RO-S3	9.255	.396	.341	36.470	13.137	11.983
ROR-HUE	21.025	.560	.312	52.977	66.406	35.330
ROR-T3	50.932	.464	.334	76.638	49.201	31.943
ROR-S3	59.006	.372	.335	81.295	17.787	15.989

R-HUE	20.776	.533	.298	52.704	65.720	25.701
R-T4	65.839	.427	.330	84.914	42.161	25.358
R-S3	7.950	.367	.330	33.879	9.231	6.926
RVR-HUE	20.994	.509	.287	52.943	64.925	18.513
RVR-S1	19.907	.394	.302	51.732	29.178	6.063
RVR-S3	56.832	.362	.319	80.085	20.698	8.946
RV-HUE	10.373	.364	.235	38.503	38.695	-12.258
RV-T2	32.298	.400	.262	63.589	54.585	-4.093
VRV-HUE	6.522	.369	.224	30.692	38.011	-12.426
VRV-S3	36.335	.402	.304	66.775	37.429	9.204
V-HUE	6.801	.277	.219	31.350	18.092	-20.326
VBV-HUE	8.199	.320	.227	34.393	27.957	-16.596
VBV-T4	50.000	.315	.277	76.069	20.125	-12.119
BV-HUE	8.540	.339	.242	35.084	27.818	-11.973
BV-S2	7.516	.303	.254	32.953	14.265	-11.708
BVB-HUE	9.876	.224	.186	37.619	16.407	-36.051
BVB-S3	47.516	.318	.317	74.518	3.003	1.323
B-HUE	15.776	.205	.207	46.680	.915	-36.717
B-T1	22.174	.209	.216	54.211	-1.318	-37.657
BGB-HUE	21.087	.202	.245	53.045	-16.690	-28.680
BGB-T3	49.689	.247	.298	75.878	-21.550	-14.150
BG-HUE	18.571	.211	.281	50.182	-24.261	-17.392
BG-T1	26.522	.218	.295	58.529	-28.883	-15.169
BG-S2	11.429	.247	.313	40.293	-16.929	-5.797
GBG-HUE	13.696	.245	.366	43.793	-30.778	3.565
GBG-S2	38.199	.283	.333	68.167	-16.869	1.158
G-HUE	20.652	.234	.379	52.567	-42.213	5.602
G-S3	8.789	.300	.348	35.574	-9.326	4.481
GYG-HUE	22.919	.251	.416	54.989	-45.709	14.888

GYG-T4	72.981	.322	.380	88.439	-21.351	22.981
GYG-S1	25.528	.312	.380	57.586	-18.204	14.968
YG-HUE	45.342	.336	.484	73.116	-41.743	49.122
YG-S3	10.590	.356	.375	38.881	-2.521	14.493
YGY-HUE	23.789	.312	.477	55.876	-39.094	34.600
YGY-S3	60.248	.359	.385	81.973	-6.990	29.444
ROSE	23.075	.493	.276	55.149	67.894	13.309
SIENNA	23.168	.519	.356	55.245	43.429	40.857
WHITE	100.000	.342	.335	100.000	6.801	13.087
GRAY-1	76.087	.341	.333	89.900	6.681	11.078
GRAY-2	60.870	.340	.333	82.309	5.781	10.122
GRAY4	40.683	.339	.332	69.953	5.063	8.426
GRAY-6	18.602	.344	.332	50.219	5.316	7.040
GRAY8	7.174	.348	.332	32.200	4.686	5.446
BLACK	5.000	.351	.335	26.735	4.129	5.467

References

Alvarado, N. (1998) A reconsideration of the structure of the emotion lexicon. *Motivation and Emotion, 22*, 329-344.

Berlin, B. & Kay, P. (1969) *Basic color terms: Their universality and evolution* Berkeley: University of California Press.

Brown, R. & Lenneberg, E. (1954) A study in language and cognition. Journal of *Abnormal and Social Psychology, 49*, 454-462.

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

Davies, I.R.L., & Corbett, G.G. (1997) A cross-cultural study of colour-grouping: Evidence for weak linguistic relativity. *British Journal of Psychology, 88*, 493-517.

Davies, I.R.L., MacDermid, C., Corbett, G.G., McGurk, H. Jerrett, D., Jerrett, T. & Sowden, P. (1992) Color terms in Setswana – A linguistic and perceptual approach. *Linguistics, 30*, 1065-1103.

Davies, I.R.L., Sowden, P., Jerrett, D.T., Jerrett, T. & Corbett, G.G. (1998) A cross-cultural study of English and Setswana speakers on a colour triads task: A test of the Sapir-Whorf hypothesis. *British Journal of Psychology, 89*, 1-15.

Davies, I. R. L. (1998). A cross-cultural study of colour-grouping: tests of the perceptual-physiology account of colour universals. *Ethos*, **26(3)**, 338-360.

Davies, I. R. L. & Corbett, G. G. (1994). Russian basic colour terms. *Linguistics,* **32**, 63-89.

Davies, I. R. L., Roling, P., Corbett, G. G., Xoagub, F. & Xoagub, J.. (1998). Color Terms and Color Term Acquisition in Damara. *The Journal of Linguistic Anthropology*, **7(2)**, 181-207.

Davies, I. R. L., Sosenskaja, T. & Corbett, G. G. (1999). First account of the basic colour terms of a Daghestanian language: the case of Tsakhur. *Linguistic Typology*, **3**,.179-207.

Dedrick, D (1998) On the foundations of the universalist tradition in colour naming (and their supposed refutation). *Philosophy of the Social Sciences 28, 179-204.*

Dye, W., Townsend, P. & Townsend, W (1966) The Sepik hill languages: A preliminary report. *Oceania, 39*, 146-156.

Gopnik, A., & Meltzoff, A. N. (1997). *Words, thoughts and theories*. Cambridge, MA: Bradford, MIT Press.

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

Ibragimov, G. Kh. (1990) *Caxurskij jazyk (The Tsakhur language)*. Moscow: Nauka.

Jameson, K.A. (1997) What Saunders and van Brakel chose to ignore in color and cognition research. *Behavioral and Brain Sciences, 20*, 195-196.

Jameson, K. A., & Alvarado, N. (2003a). Differences in color naming and color salience in Vietnamese and English. *Color Research & Application, 28*, 113-138.

Jameson, K. A. & Alvarado, N. (2003b) The relational correspondence between category exemplars and names. *Philosophical Psychology, 16*, 26-49.

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

Kibrik, A. E. & S. V. Kodzasov, (1990). *Sopostavitel´noe izu_enie dagestankix jazykov: imja, fonetika.* [Comparative Study of the Daghestanian Languages: Nominals, Phonetics.] Moscow: Izdatel´stvo Moskovskogo universiteta.

Kibrik, A. E. (ed.) (1999). *Èlementy caxurskogo jazyka v tipologi_eskom osve__enii*. [Elements of Tsakhur from a Typological Perspective.] Moscow: Nasledie Press.

Kruskal, J.B. & Wish, M. (1981) *Multidimensional scaling*. London: Sage Publications.

Levinson, S.C., Kita, S., Haun, D.B.M. & Rasch, B.H. (2002) Returning the tables: language affects spatial reasoning. *Cognition, 84,* 155-188.

MacLaury, R.E. (1997) Color and cognition in mesoAmerica: *Constructing categories as vantages*. Austin, Texas: University of Texas Press.

Malan, J.S. (1995) *Peoples of Namibia*. Pretoria, SA: Rhino Publishers

Malt, B.C. (1995) Category coherence in cross-cultural perspective. *Cognitive Psychology 29*, 85-148.

Mervis, C.B. & Bertrand, J. (1994) Acquisition of the novel name nameless category (N3C) principle. *Child Development, 65*, 1646-1662.

Moss, A., Davies, I., Corbett, G. & Laws, G. (1990) Mapping Russian basic color terms using behavioral measures. *Lingua, 82*, 313-332.

Munnich, E., & Landau, B. (2003). The effects of spatial language on spatial representation: Setting some boundaries. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and thought*. Cambridge, MA: MIT Press.

Norusis, M.J. (1994). *SPSS Professional Statistics TM 6.1*. Chicago: SPSS Inc.

Özgen, E. & Davies , I.R.L (1998) Turkish color terms: tests of Berlin and Kay's theory of color universals and linguistic relativity. *Linguistics, 36*, 919-956.

Paramei, G.V. (2005) Singing the Russian blues: An argument for culturally determined basic color terms. *Cross-Cultural Research, 39*, 10-38.

Pilling, M. & Davies, I.R.L. (2004) Linguistic relativism and colour cognition. *British Journal of Psychology, 95,* 429-455.

Roberson, D., Davidoff, J. & Braisby, N. (1999) Similarity and categorisation: Neuropsychological evidence for a dissociation in explicit categorisation tasks. *Cognition, 71*, 1-42.

Roberson, D., Davidoff, J., Davies, I.R.L. & Shapiro, L. R. (2004) The Development of Color Categories in Two languages: a longitudinal study. *Journal of Experimental Psychology: General, 133*, 554-571.

Roberson, D., Davidoff, J., Davies, I. & Shapiro, L. Colour categories in Himba: Evidence for the cultural relativity hypothesis. *Cognitive Psychology (in press)*

Roberson, D., Davies I. & Davidoff, J. (2000) Colour categories are not universal: Replications and new evidence from a Stone-age culture. *Journal of Experimental Psychology: General , 129*, 369-398.

Rosch, E.H. (1973) Natural categories. *Cognitive Psychology, 4,* 328-350.

Rosch Heider, E. (1972) Universals in color naming and memory. *Journal of Experimental Psychology, 93*, 10-20.

Rosch, E. (1975) Cognitive representations of semantic categories. *Journal of Experimental Psychology: General 104*, 192-253.

Saunders, B.A.C. & van Brakel, J. (1997) Are there non-trivial constraints on colour categorization . *Behavioral & Brain Sciences, 20*, 167-178.

Schyns, P.G., Goldstone, R.L. & Thibaut, J.P. (1998) The development of features in object concepts, *Behavioral and Brain Sciences, 21*, 1-56.

Shaver, P., Schwartz, J., Kirson, D. & O'Connor, C. (1987) Emotion knowledge: Further exploration of a prototype approach. *Journal of Personality and Social Psychology, 52,* 1061-1086.

Steels L. & Belpaeme, T. Coordinating Perceptually Grounded Categories through Language. A Case Study for Colour. *Behavioral and Brain Sciences, (in press).*

Wierzbicka, A. (1990) The meaning of color terms: semantics, culture and cognition. *Cognitive Linguistics, 1*, 99-150.

Table 1. Number of informants, male / female ratio and mean age for each language group.

Language	Total N	Males / Females	Mean age
English (1)	18	10 / 8	22
English (2)	47	24 / 23	29
Russian	77	24 / 53	34
Bagwalal	25	14 / 11	36.8
Tsakhur	19	11 / 8	36.4
Turkish	34	15 / 19	29
Berinmo	17	1 / 16	34
Damara	40	15 / 25	35.1
Nama	56	23 / 33	36.6
Ndonga	80	37 / 43	30.0
Herero	20	10 / 10	34.4
Himba	21	1 / 20	34.2
Kwanyama	20	7 / 13	28.1
Kwangali	15	10 / 5	28.1
Mbukushu	10	5 / 5	30.8
Shona	39	20 / 19	37
Tswana	44	22 / 22	45
!Xoo	7	4 / 5	30

Table 2. Mean number of groups, 95% confidence limits and number of basic color terms for each language

Language	Sample size	number of groups	Confidence limits ± 1.96 SE	Number of basic terms
Bagwalal	25	33.6	4.07	7
Berinmo	17	20.3	0.69	5
Damara	40	10.5	0.43	11
English1	18	13.7	3.66	11
Herero	20	16.4	2.94	11
Kwanyama	20	11.6	1.35	8
Nama	56	15.7	1.82	11
Ndonga	80	13.5	1.10	11
Kwangali	15	11.7	1.78	6
Russian	75	13.5	1.72	12
Mbukushu	10	10.1	2.25	6
Tsakhur	19	11.3	2.16	11
Tswana	44	17.3	1.92	6
Shona	17	12.4	1.61	6
Turkish	23	13.2	2.78	12
English2	47	12.4	1.92	11
!Xoo	7	15.4	5.21	5
Himba	21	21.4	4.10	5

Figure captions:

Figure 1a. The 65 Color Aid tiles used in the free-sorting task plotted in the a* x b* dimensions of CIELab space.

Figure 1b. The 65 Color Aid tiles used in the free-sorting task plotted in the a* x L* dimensions of CIELab space.

Figure 1c. The 65 Color Aid tiles used in the free-sorting task plotted in dimensions 2 x 1 dimensions of the MDS dimensions of best fit.

Figure 1d. The 65 Color Aid tiles used in the free-sorting task plotted in dimensions 1 x 3 dimensions of the MDS dimensions of best fit.

Figure 1e. Mean sorting responses for the 18 languages tested in the free-sorting task weighted on dimensions 1 x 2 of the MDS dimensions of best fit.

Figure 1f. Mean sorting responses for the 18 languages tested in the free-sorting task weighted on dimensions 1 x 3 of the MDS dimensions of best fit.

Figure 2a. The 21 Color Aid tiles in the (English) purple-blue-green categories plotted in the a* x b* dimensions of CIELab space.

Figure 2b. The 21 Color Aid tiles in the (English) purple-blue-green categories plotted in the a* x L* dimensions of CIELab space.

Figure 2c. The 21 Color Aid tiles in the (English) purple-blue-green categories plotted in dimensions 2 x 1 dimensions of the MDS dimensions of best fit.

Figure 2d. The 21 Color Aid tiles in the (English) purple-blue-green categories plotted in dimensions 2 x 3 dimensions of the MDS dimensions of best fit.

Figure 2e. Mean sorting responses for the 18 languages tested for the (English) purple-blue-green categories weighted on dimensions 2 x 1 of the MDS dimensions of best fit.

Figure 2f. Mean sorting responses for the 18 languages tested for the (English) purple-blue-green categories weighted on dimensions 2 x 3 of the MDS dimensions of best fit.

Figure 3a. The 14 Color Aid tiles in the (English) purple-pink categories plotted in the a* x b* dimensions of CIE Lab space.

Figure 3b. The 14 Color Aid tiles in the (English) purple-pink categories plotted in the a* x L* dimensions of CIE Lab space.

Figure 3c. The 14 Color Aid tiles in the (English) purple-pink categories plotted in dimensions 2 x 1 dimensions of the MDS dimensions of best fit.

Figure 3d. The 14 Color Aid tiles in the (English) purple-pink categories plotted in dimensions 2 x 3 dimensions of the MDS dimensions of best fit.

Figure 3e. Mean sorting responses for the 18 languages tested for the (English) purple-pink categories weighted on dimensions 2 x 1 of the MDS dimensions of best fit.

Figure 3f. Mean sorting responses for the 18 languages tested for the (English) purple-pink categories weighted on dimensions 2 x 3 of the MDS dimensions of best fit.

Figure 4a. The 12 Color Aid tiles in the (English) red-orange-pink categories plotted in the a* x b* dimensions of CIELab space.

Figure 4b. The 12 Color Aid tiles in the (English) red-orange-pink categories plotted in the a* x L* dimensions of CIELab space.

Figure 4c. The 12 Color Aid tiles in the (English) red-orange-pink categories plotted in dimensions 2 x 1 dimensions of the MDS dimensions of best fit.

Figure 4d. The 12 Color Aid tiles in the (English) red-orange-pink categories plotted in dimensions 3 x 1 dimensions of the MDS dimensions of best fit.

Figure 4e. Mean sorting responses for the 18 languages tested for the (English) red-orange-pink categories weighted on dimensions 2 x 1 of the MDS dimensions of best fit.

Figure 4f. Mean sorting responses for the 18 languages tested for the (English) red-orange-pink categories weighted on dimensions 3 x 1 of the MDS dimensions of best fit.

Figure 1a

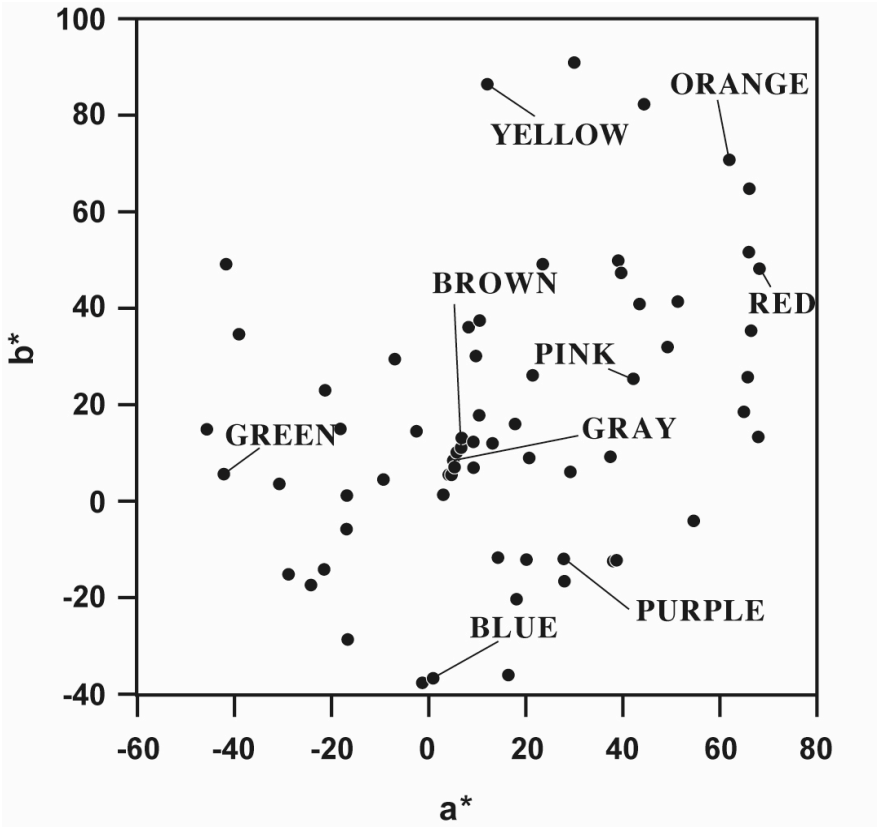

100
80
60
40
20
0
-20
-40
b*
-60
-40
-20
0
20
40
60
80
a*
YELLOW
ORANGE
BROWN
RED
PINK
GREEN
GRAY
PURPLE
BLUE

Figure 1b

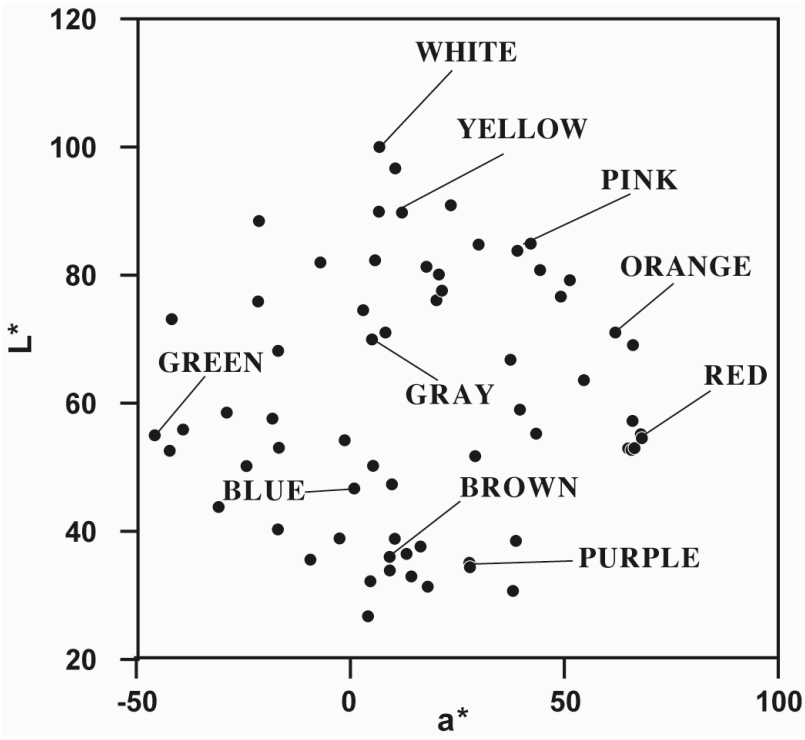

WHITE
YELLOW
PINK
ORANGE
GREEN
GRAY
RED
BLUE
BROWN
PURPLE
L*
a*
120
100
80
60
40
20
-50
0
50
100

Figure 1c

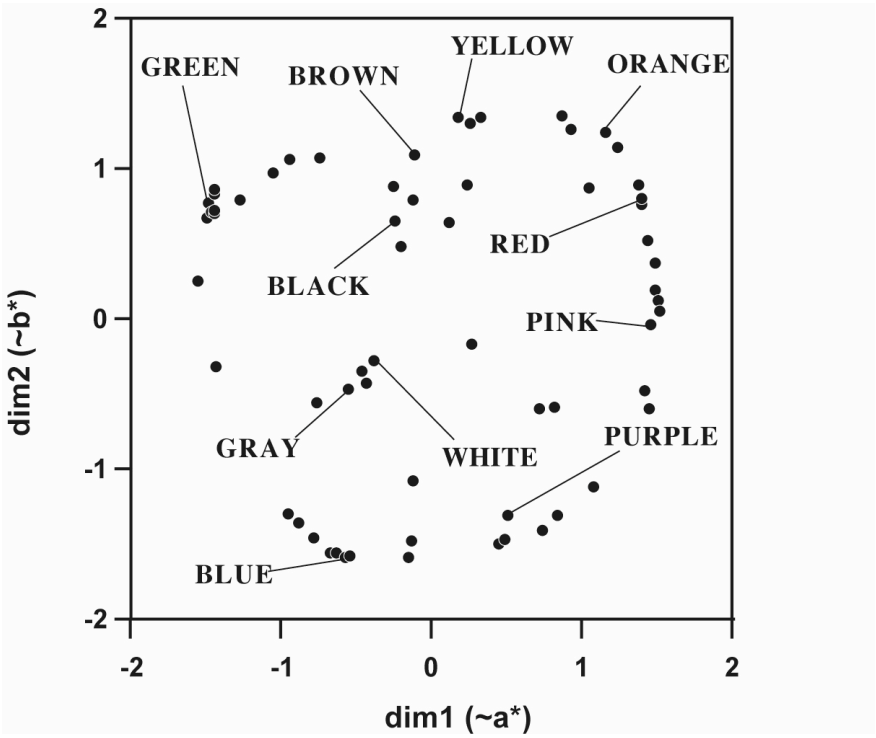

GREEN
BROWN
YELLOW
ORANGE
RED
BLACK
PINK
GRAY
WHITE
PURPLE
BLUE
dim2 (~b*)
dim1 (~a*)
2
1
0
-1
-2
-2
-1
0
1
2

Figure 1d

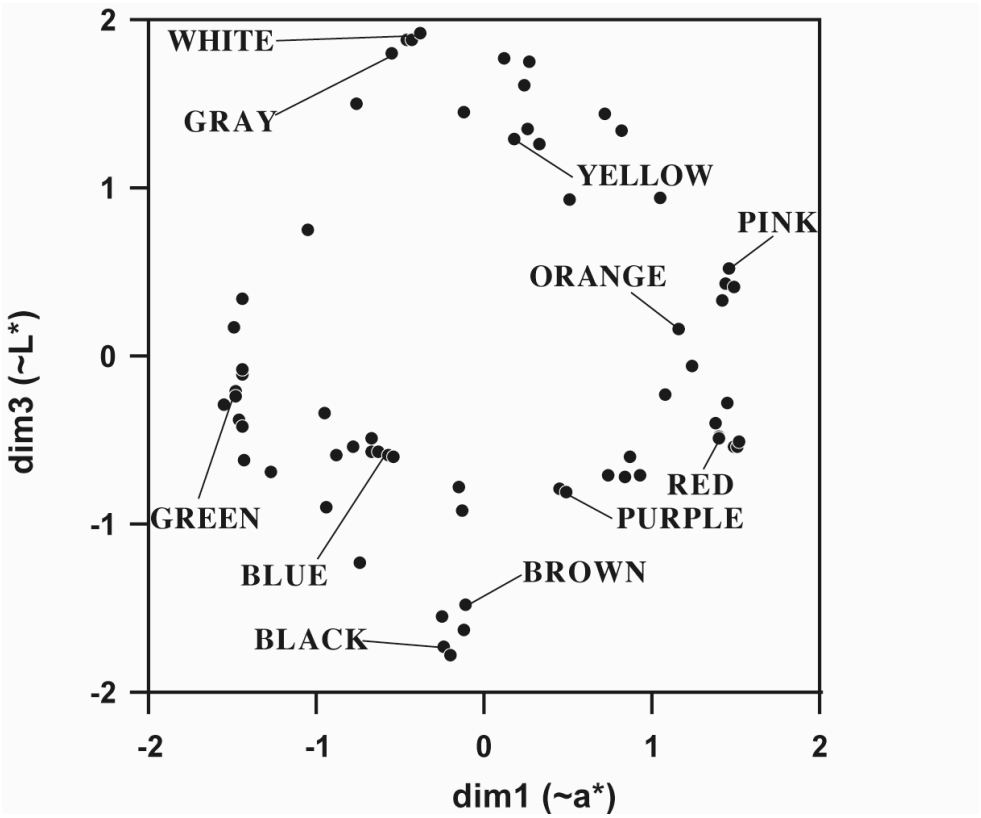

WHITE
GRAY
YELLOW
PINK
ORANGE
RED
PURPLE
GREEN
BLUE
BROWN
BLACK
dim3 (~L*)
dim1 (~a*)
2
1
0
-1
-2
-2
-1
0
1
2

Figure 1e

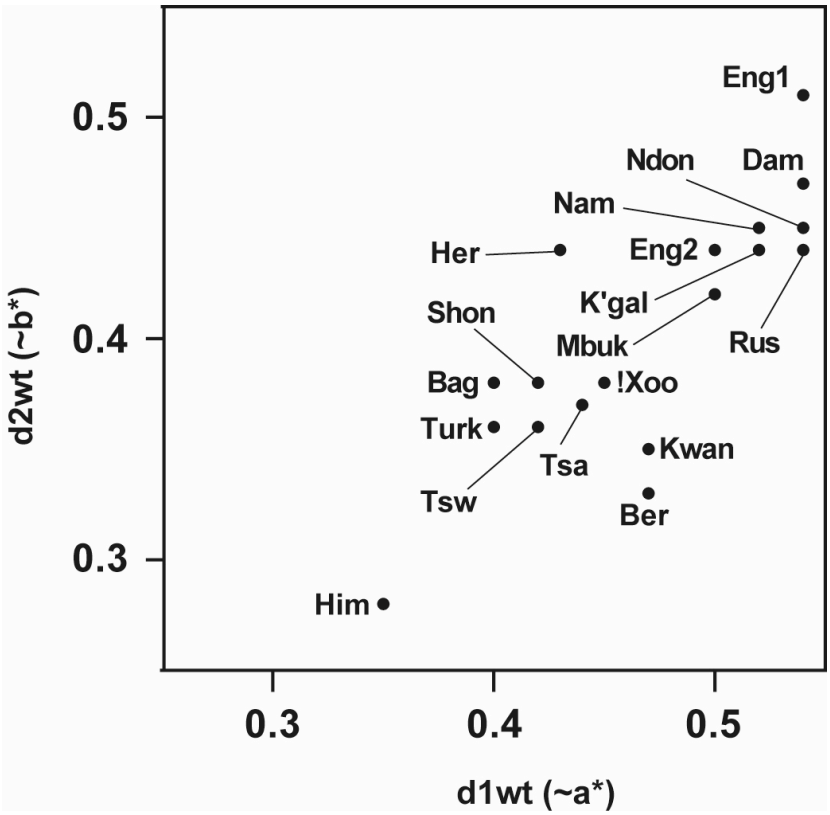

Eng1
Ndon
Dam
Nam
Her
Eng2
K'gal
Shon
Mbuk
Rus
Bag
!Xoo
Turk
Kwan
Tsa
Tsw
Ber
Him
0.5
0.4
0.3
d2wt (~b*)
0.3
0.4
0.5
d1wt (~a*)

Figure 1f

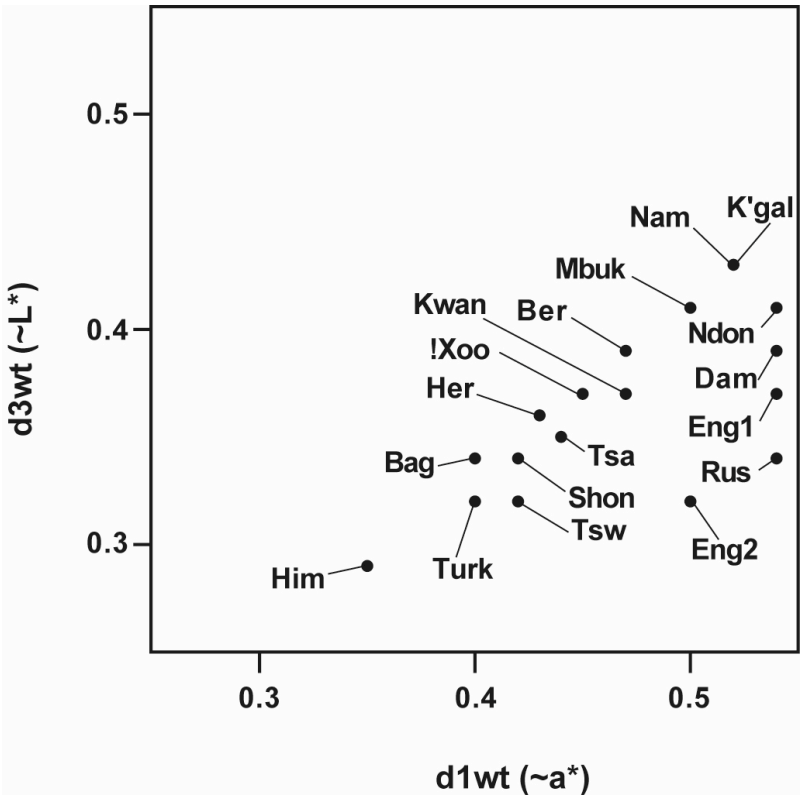

0.5
0.4
0.3
d3wt (~L*)
0.3
0.4
0.5
d1wt (~a*)
Nam
K'gal
Mbuk
Kwan
Ber
Ndon
!Xoo
Dam
Her
Eng1
Tsa
Bag
Rus
Shon
Tsw
Eng2
Him
Turk

Figure 2a

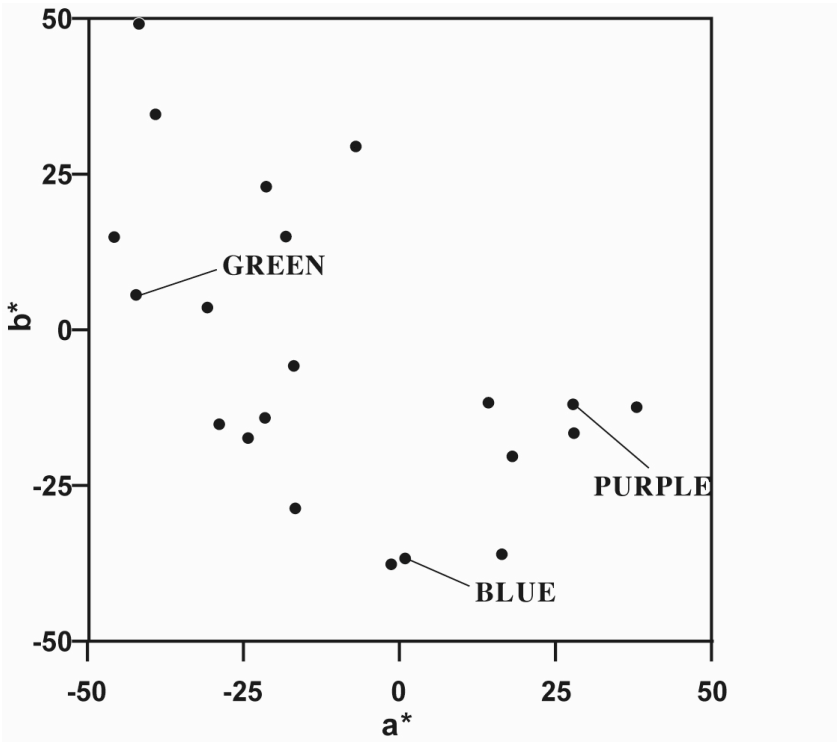


Figure 2b

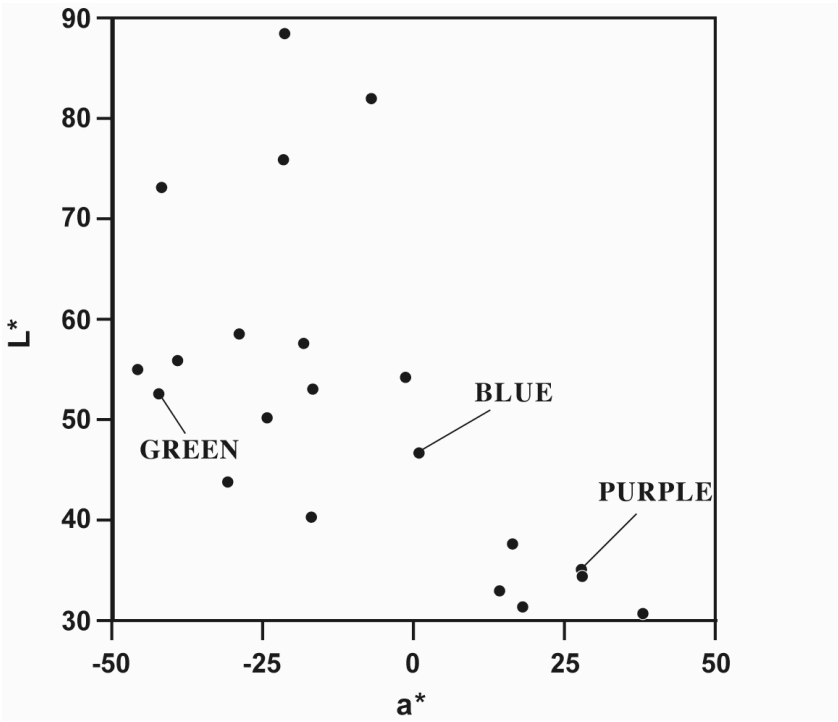


Figure 2c

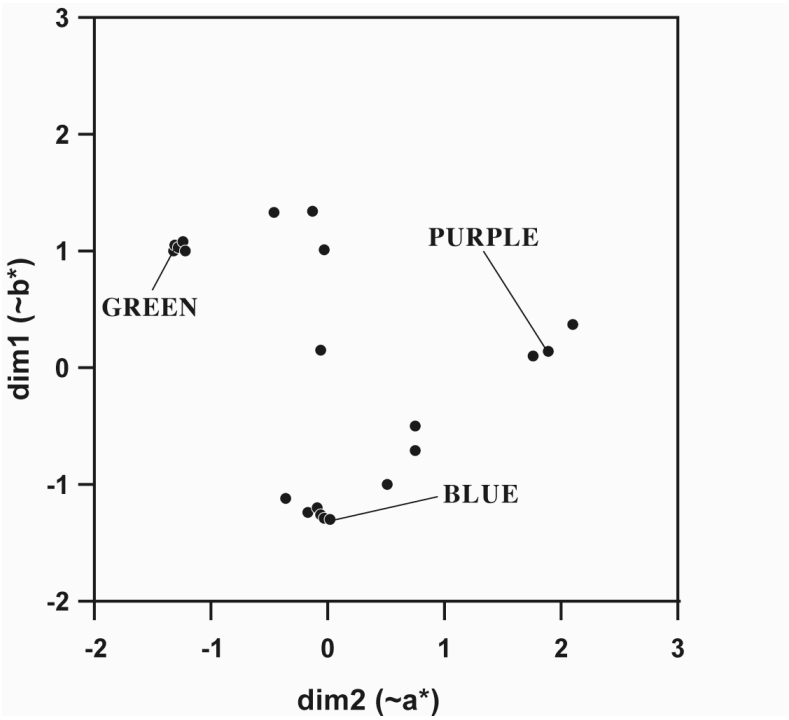


Figure 2d

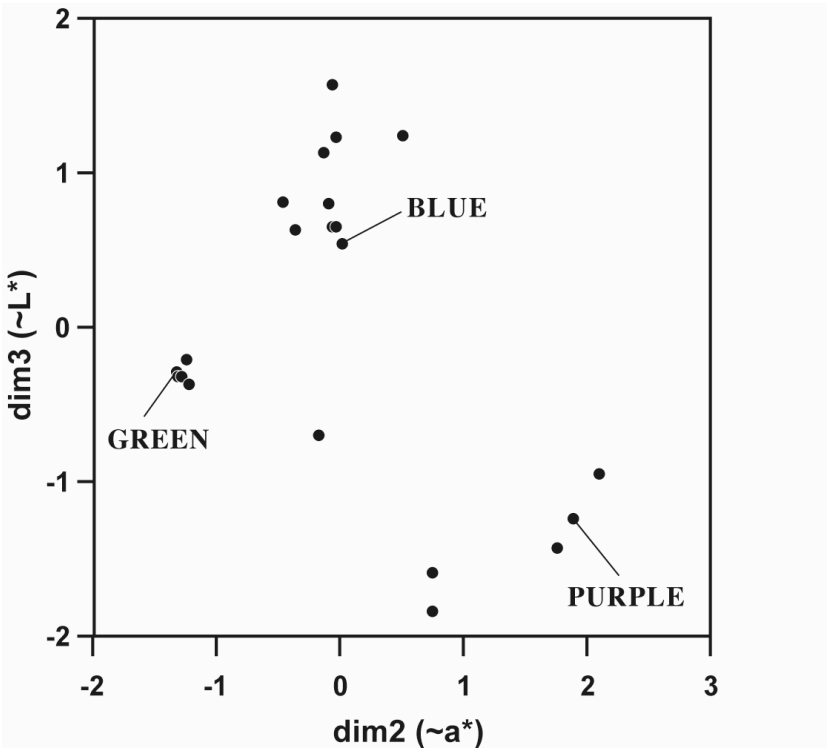


Figure 2e

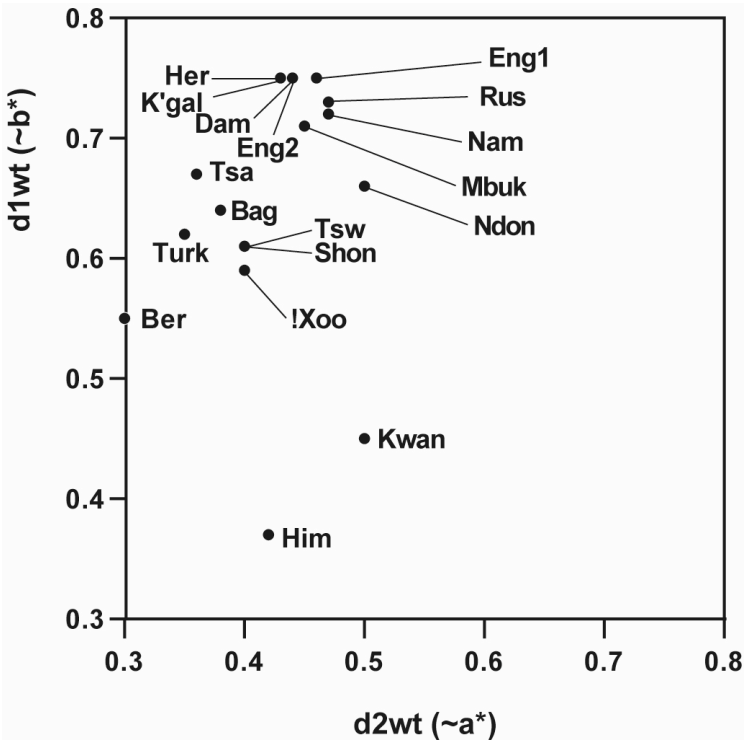

0.8
0.7
0.6
0.5
0.4
0.3
d1wt (~b*)
0.3
0.4
0.5
0.6
0.7
0.8
d2wt (~a*)
Her
K'gal
Dam
Eng2
Eng1
Rus
Nam
Mbuk
Ndon
Tsa
Bag
Tsw
Shon
Turk
!Xoo
Ber
Kwan
Him

Figure 2f

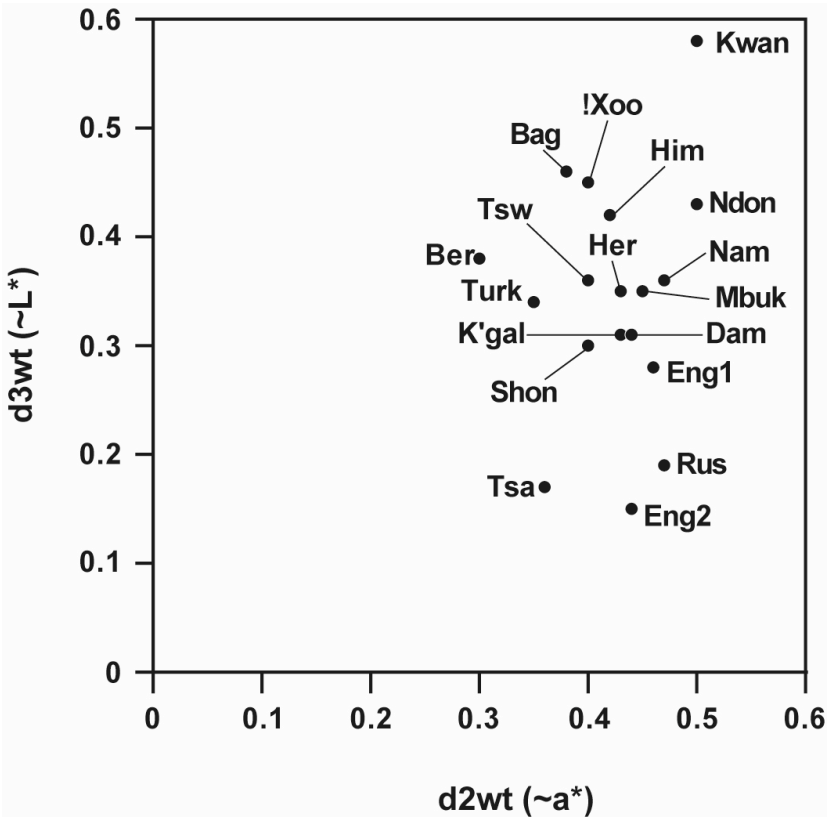

0.6
0.5
0.4
0.3
0.2
0.1
0
d3wt (~L*)
0
0.1
0.2
0.3
0.4
0.5
0.6
d2wt (~a*)
Kwan
!Xoo
Bag
Him
Ndon
Tsw
Her
Nam
Ber
Turk
Mbuk
K'gal
Dam
Eng1
Shon
Rus
Tsa
Eng2

Figure 3a

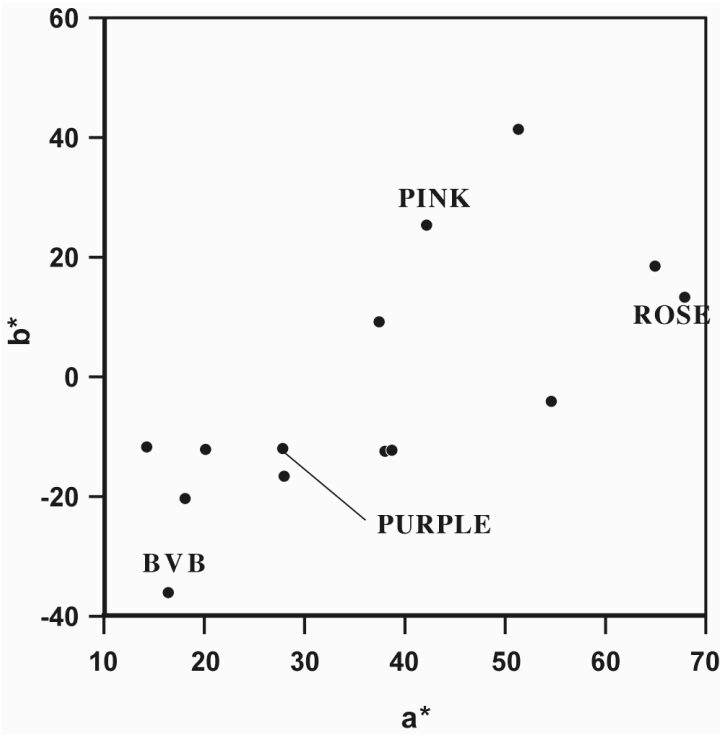


Figure 3b

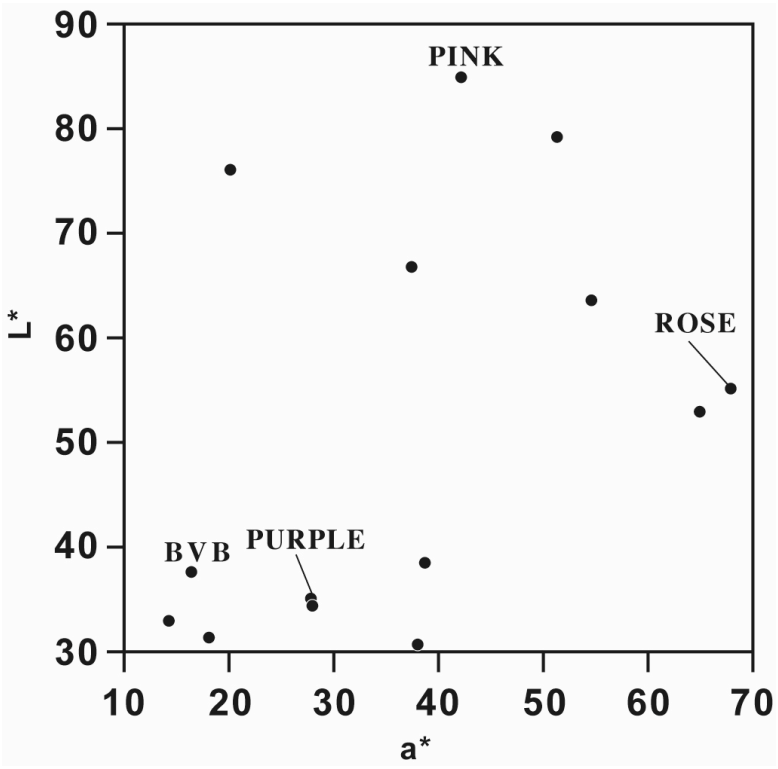


Figure 3c

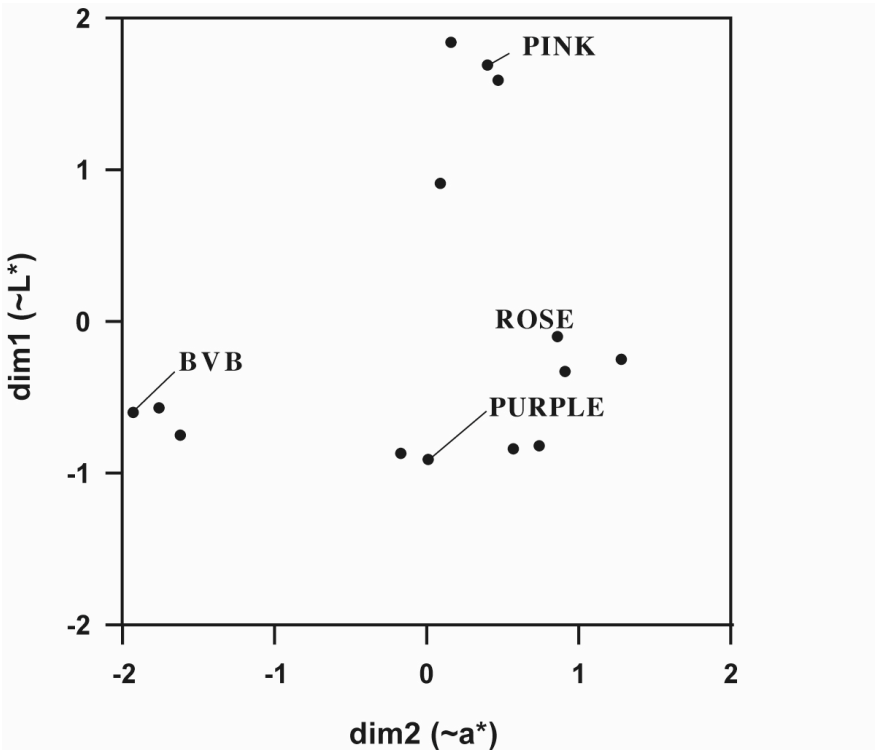


Figure 3d

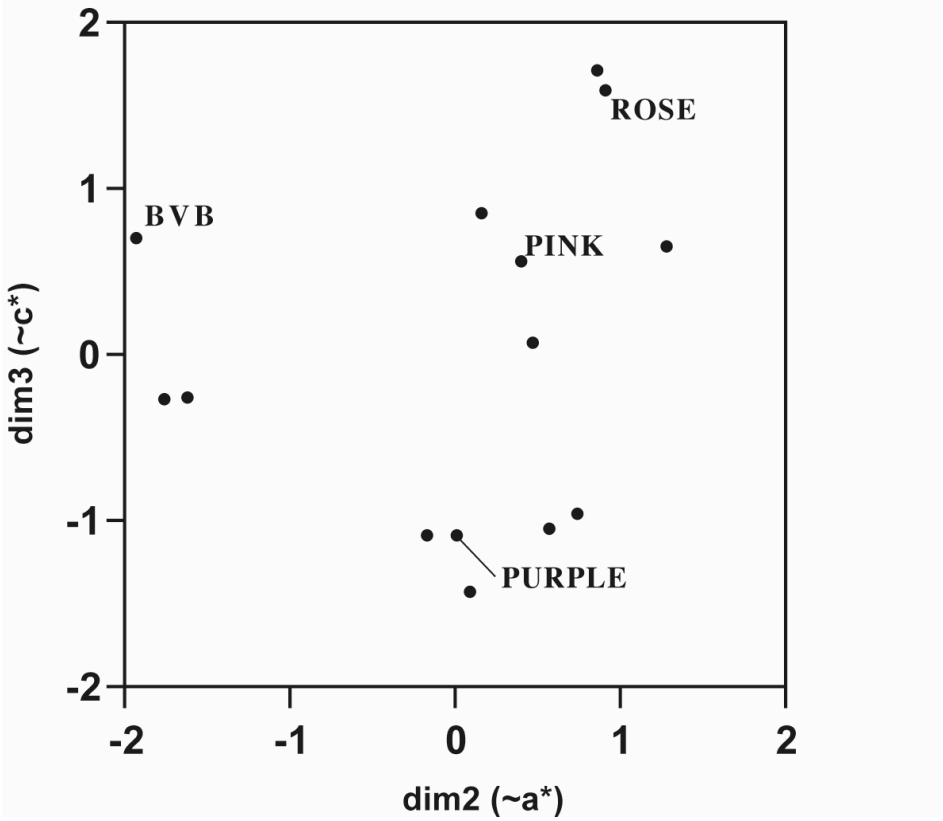


Figure 3e

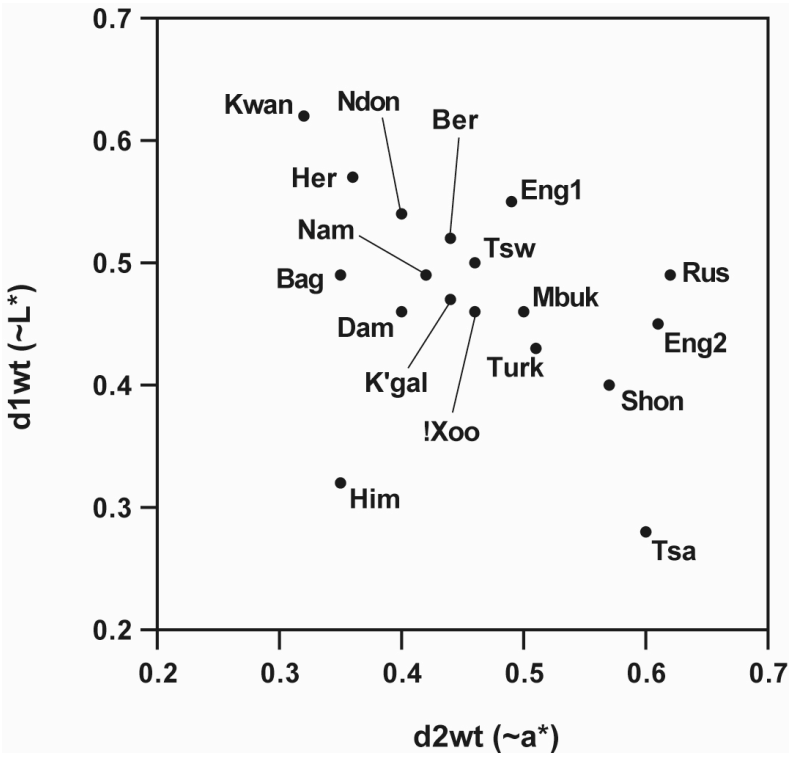

0.7
0.6
0.5
0.4
0.3
0.2
d1wt (~L*)
0.2
0.3
0.4
0.5
0.6
0.7
d2wt (~a*)
Kwan
Ndon
Ber
Her
Eng1
Nam
Tsw
Bag
Rus
Mbuk
Dam
Eng2
Turk
K'gal
Shon
!Xoo
Him
Tsa

Figure 3f

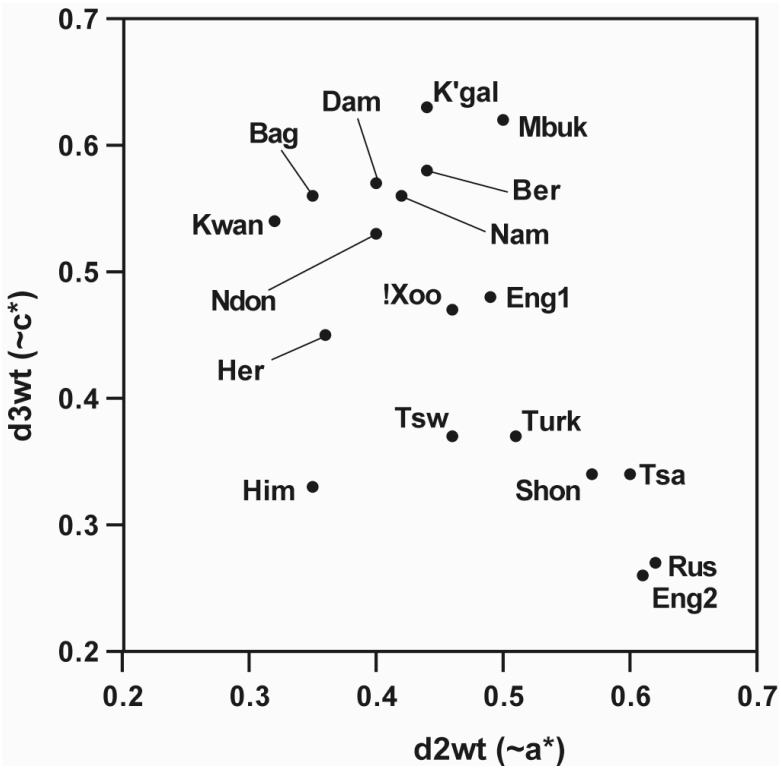

0.7
0.6
0.5
0.4
0.3
0.2
d3wt (~c*)
0.2
0.3
0.4
0.5
0.6
0.7
d2wt (~a*)
Dam
K'gal
Mbuk
Bag
Ber
Kwan
Nam
Ndon
!Xoo
Eng1
Her
Tsw
Turk
Him
Shon
Tsa
Rus
Eng2

Figure 4a

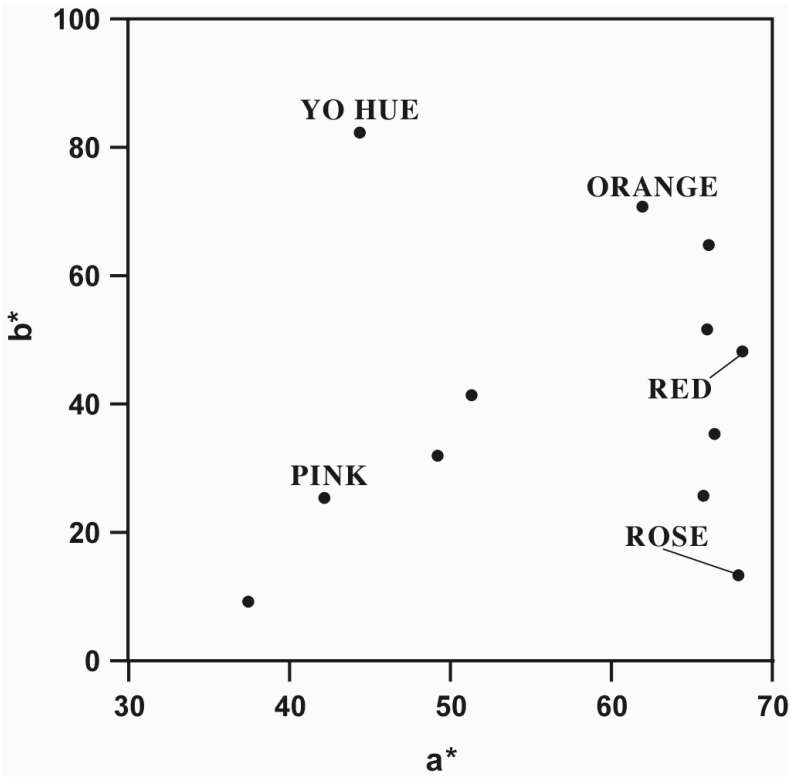


Figure 4b

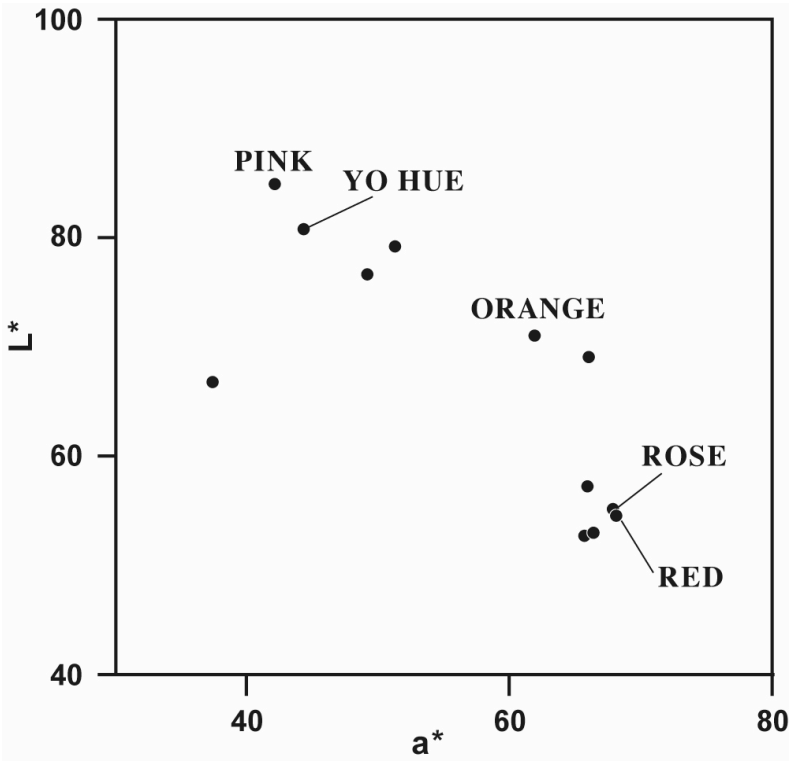


Figure 4c

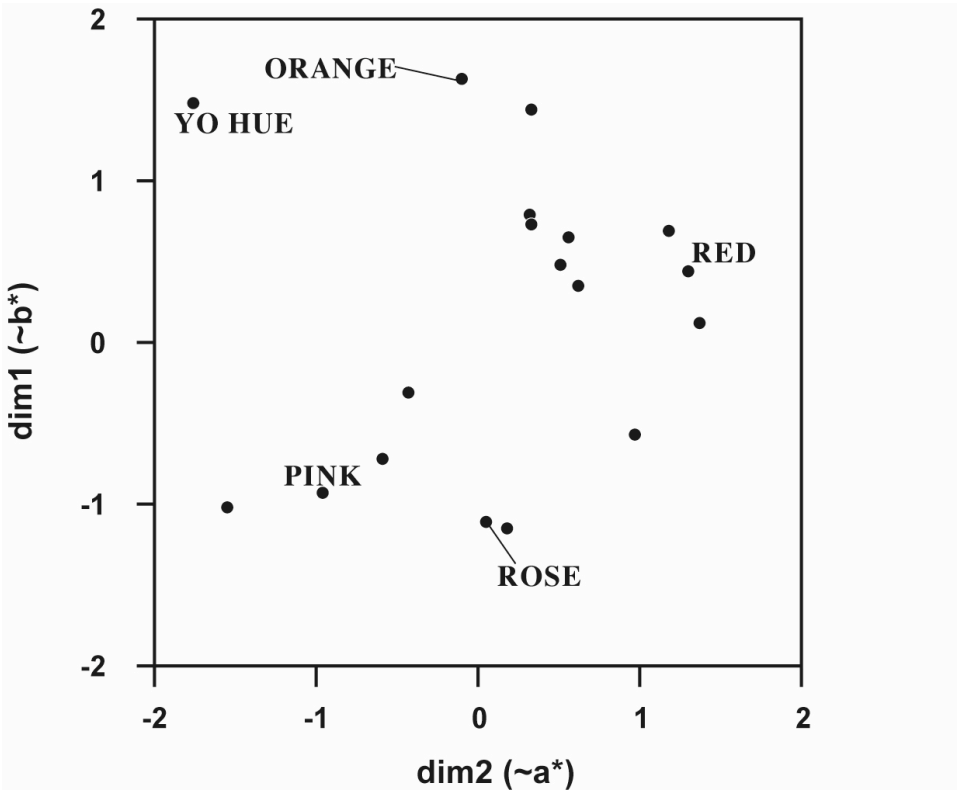


Figure 4d

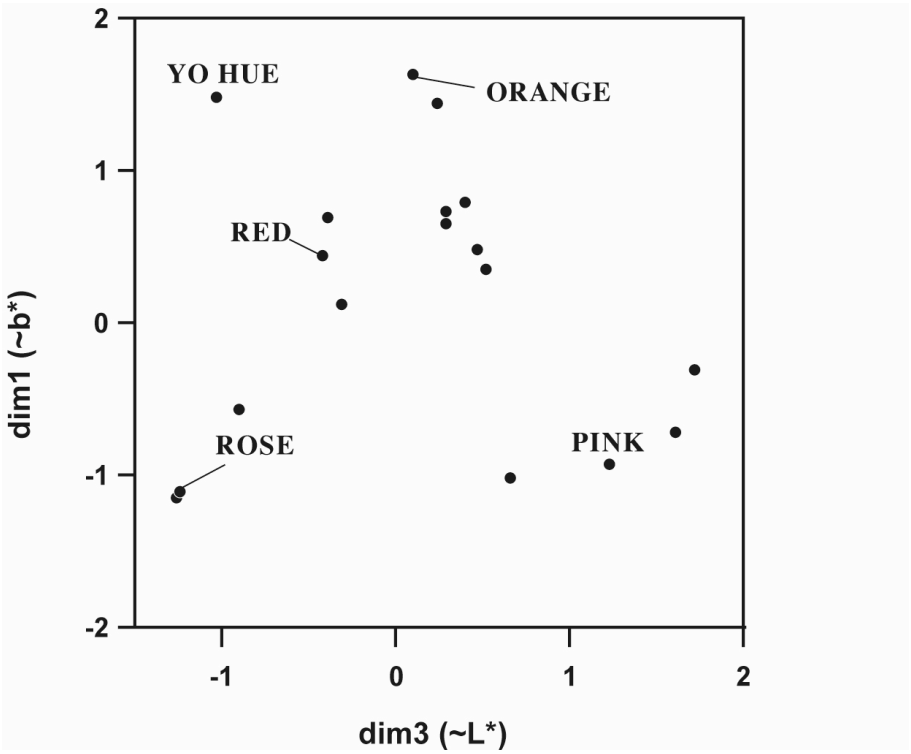


Figure 4e

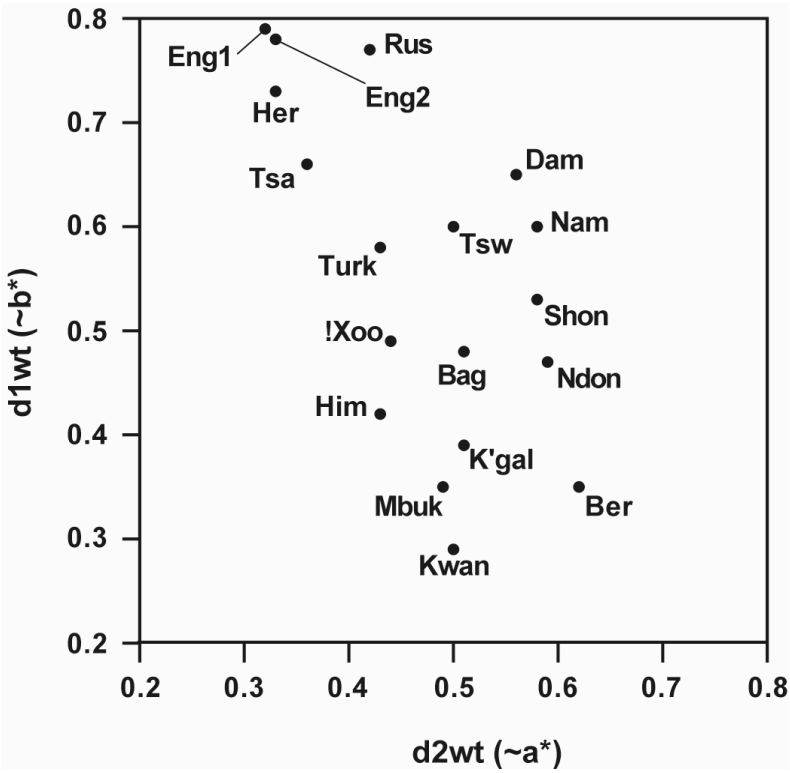
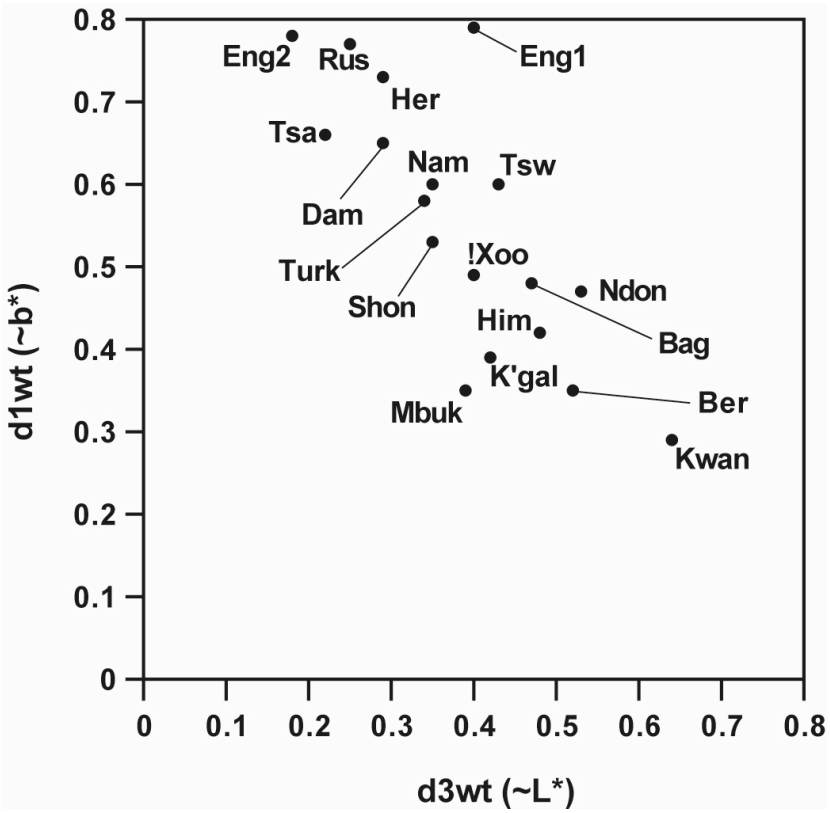

0.8
0.7
0.6
0.5
0.4
0.3
0.2
d1wt (~b*)
0.2
0.3
0.4
0.5
0.6
0.7
0.8
d2wt (~a*)
Eng1
Rus
Eng2
Her
Tsa
Dam
Nam
Tsw
Turk
Shon
!Xoo
Bag
Ndon
Him
K'gal
Mbuk
Ber
Kwan

Figure 4f


0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
d1wt (~b*)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
d3wt (~L*)
Eng2
Rus
Eng1
Her
Tsa
Nam
Tsw
Dam
Turk
!Xoo
Shon
Ndon
Him
Bag
K'gal
Ber
Mbuk
Kwan