

A cross-cultural study of colour grouping: Evidence for weak linguistic relativity

Ian R. L. Davies* and Greville G. Corbett

Department of Psychology, University of Surrey, Guildford, Surrey GU2 5XH, UK

We report a cross-cultural study of colour grouping carried out as a test of the Sapir–Whorf hypothesis (linguistic relativity theory). Speakers of English, Russian and Setswana—languages that differ in their number of basic colour terms, and in how the blue–green region is categorized—were compared on a colour sorting task. Informants sorted a representative set of 65 colours into groups so that members of the groups looked similar to each other, with no restriction on the number of groups formed. If linguistic relativity theory is true, then there should be reliable differences between the three samples in the composition of the groups they formed associated with the differing positions of colour category boundaries in the languages. The most striking feature of the results, inconsistent with linguistic relativity theory, was the similarity amongst the patterns of choice of the three samples. However, there were also significant differences amongst the samples. Setswana speakers (who have a single basic term for BLUE or GREEN) were more likely to group BLUE colours with GREEN colours than either English or Russian speakers. But Russian speakers (who have two basic colour terms for BLUE) were no more likely than English speakers to group light and dark BLUE separately. In addition there were general structural differences in grouping among the samples: they differed in the level of consensus in grouping, the number of groups formed and in the distribution of the number of colours placed in a group. These structural differences may reflect differences in the availability and salience of the colour categories across the languages. Our data support perceptual universalism modulated by weaker linguistic effects.

Why do we see the world the way we do (Koffka, 1935)? Perhaps the ‘orthodox’ answer to Koffka’s question for most of this century is that our minds impose structure on sensations, and thus appearances are the result of mental construction. The major alternative to constructionism is the theory of direct perception usually attributed to J. J. Gibson (see Fodor & Pylyshyn, 1981; Turvey, 1977). Linguistic relativity theory (Whorf, 1956) is allied to constructionism. Proponents of Whorfianism maintain that language is a major vector of constructionism: the language we speak influences the way we think and perceive, and differences in grammar and vocabulary across languages reveal differences in cognition. We report here a cross-cultural study of colour grouping carried out in part as a test of the Whorfian thesis.

* Requests for reprints.

Colour cognition has been the natural testing ground for the Whorfian hypothesis.¹ It has been known since at least the time of the Cambridge expedition to the Torres Straits (Rivers, 1901) that the size of colour-term inventories varies across languages. Some languages, such as that spoken by the Dani of New Guinea, have just two 'basic' (frequent, salient, high consensus) colour terms (Heider, 1972); other languages, including most southern African languages, have four or five basic colour terms; whereas other languages (including English) have 11 basic colour terms. Such differences were embraced by the constructionists and cultural relativists as clear examples of the arbitrariness and relativity of perceptual categories (e.g. Gleason, 1961); the segmentation of the physical continuum of the spectrum into colour categories was held to be determined by local need, rather than universal influences, such as common perceptual physiology.²

As Brown (1976) and Lucy & Shweder (1979) point out, work within the colour framework divides into two phases. Until about 1970 the prevailing assumption was that the physical continuum of the spectrum was segmented arbitrarily into categories, corresponding to the lexical terms found in language. During this phase, experimental tests of the Whorfian hypothesis of linguistic influences on colour cognition typically used recognition memory as the cognitive task, and investigated whether recognition is affected either by 'codability' or 'communication accuracy' (e.g. Brown & Lenneberg, 1954; Lantz & Steffle, 1964). The relativistic premise was so engrained that most experiments just studied speakers of a single language (usually English). Thus, for example, Brown & Lenneberg's (1954) classic study of the relationship between codability and recognition memory for colours was carried out only on speakers of English, because it was assumed that if codability affected recognition, then people with different colour lexicons would necessarily have different colour codabilities, and therefore different colour recognition memory. It is of course possible that perceptual distinctiveness determines codability, and that perceptual distinctiveness could be universal rather than culturally relative.

In the second phase of research on language and colour, the prevailing belief shifted from cultural relativism towards belief in colour universals. Two crucial studies provided much of the impetus for the shift towards universalism: Berlin & Kay's (1969) study of the relative positions in colour space of the referents of colour terms from 98 languages and Heider's (1972) study of the Dani of New Guinea.³ Berlin & Kay (1969) suggested that there was much less variation in the foci (the best examples) of colour terms across languages than there was in the boundaries of

¹ See, however, Levinson & Brown's (1994) brilliant discussion of the universality of spatial concepts.

² There have also been suggestions that the language differences reflect inherited differences in perception. Rivers (1901) reported that the inhabitants of the Torres Straits had 'weak' perception of blue. Bornstein (1973) argued that there would be a survival advantage for people living in regions with high ultra-violet (UV) irradiation, such as the tropics, to have a short wavelength filter, in order to reduce the amount of UV incident on the retinae. Such a filter would reduce the retinal damage produced by UV, but it would also reduce sensitivity to visible short wavelengths (blue and green). We have found (Davies, Laws, Jerrett & Corbett, submitted) some evidence that people in rural Africa are mildly 'tritanopic' (blue-yellow colour vision anomaly), but it is likely that this is due to 'premature ageing' of the lens due to the cumulative effect of UV.

³ See Lucy (1992), Ratner (1989) and Simpson (1991) for critical discussions of Berlin and Kay (1969) and Heider (1972) respectively.

colour terms—the regions where one category changed to another. They claimed that all the languages they studied had colour terms whose foci were drawn from a common inventory of just 11 ‘universal’ foci illustrated in the series:

[BLACK–WHITE] → RED → [GREEN–YELLOW] → BLUE → BROWN →
[PURPLE–PINK–ORANGE–GREY]

Categories grouped in brackets share a common position in the series; thus, there are just six positions shared by the eleven categories (we use small capitals to denote Berlin & Kays’s proposed universal categories whereas colour terms denoting these categories in particular languages are given in *italic*). Further, there is an implicational structure to the series: if a language has a term in a given position in the series, then according to the theory, it should have all the terms with earlier positions in the series: so, for instance, if a language has a term for BROWN, then it should also have terms for BLUE, YELLOW, GREEN, RED, BLACK and WHITE.

Kay & McDaniel (1978) proposed that the origin of colour universals lay in universal visual physiology and perception. The foci of the first six terms in the series—the ‘primary’ terms—were more perceptually salient than the foci of the remaining terms, the ‘derived’ terms. The primary terms corresponded to the colours of Hering’s (1920) opponent process theory of colour vision, and these colours seemed to be physiologically ‘basic’, at least at the level of the mid-brain (see, for example, Jameson, 1985; but see also De Valois & De Valois, 1993, for an outline of problems with the model).

Colour categories in languages with relatively few colour terms often encompass larger areas of colour space than colour categories in languages with more extensive basic colour term inventories: all of colour space is shared between the inventory of basic terms. Thus, for instance, Setswana (the main language of Botswana) has a basic term *botala* ‘green or blue’ (henceforth denoted by the shorthand ‘grue’) which includes GREEN and BLUE (Davies *et al.*, 1992). If linguistic relativity is true, then instances of *botala* ‘grue’ should be treated as equivalent for most purposes and distinctions amongst them should be reduced relative to speakers of languages that encoded the region with more than one term.

Heider (1972) studied learning and memory for colours in the Dani of New Guinea whose language has just two basic colour terms: *mili* ‘light’ and *mola* ‘dark’. She found (Expt 3) that the Dani remembered focal colours better than non-focal colours, although they had no basic terms to describe them. Further, the Dani learned to name focal examples of the ‘missing’ terms more readily than non-focal colours (Expt 4). Heider interpreted her results as strong support for the universal perceptual salience of the focal colours.

Further support for the universal perceptual salience of the focal colours comes from Turton’s (1980) report on the colour terms of the Mursi—a people from a remote part of Ethiopia. Mursi general colour terms are based on their cattle colour naming system. For instance, *golonyi* ‘red’ is based on the cattle colour ‘reddish-brown’. Despite the origin of the term, when informants were presented with a ‘highly saturated or “pure” red stimulus’ (p. 327) they tended to add an intensifying suffix *goloin-tul* ‘really red’. Turton reports that because of their geographical position

and cultural isolation, this was probably the first time that his informants had seen such artificial stimuli, and yet they elicited such strong responses.

Although Berlin & Kay were influential in shifting the balance of the relativist–universalist argument in the direction of universalism, they did not rule out linguistic relativity: languages did differ in the number of basic terms, and universals were universals of foci, rather than of boundaries. In principle, if languages differed in either of these two factors, there was scope for the linguistic differences to produce corresponding differences in behaviour. Similarly, although Heider's work has been enormously influential, her results leave open the possibility that language affects colour memory, in addition to providing evidence consistent with universal perceptual salience of the focal colours. As Ratner (1989) points out, although the Dani remembered focal colours better than non-focal colours, they were considerably worse overall than their American comparison group. This could have been due to the Dani being less used to memory 'games' than the Americans, as Heider suggested, but it could also have been due to their restricted colour vocabulary.

Lucy & Shweder (1979) argued that Heider's results did not support the universalist position at all: rather, the results were an artifact of the (unintended) distinctiveness of the focal colours in the particular array of Munsell colours that she used. Lucy and Shweder constructed an alternative array in which they attempted to make all colours equally distinctive. They found that participants were not significantly better at recognizing focal colours than non-focal colours in the short-term memory task (although the mean scores across participants for the focal colours were about 10 per cent greater for the focal colours), whereas recognition of the focal colours was significantly better than the non-focal colours in the long-term memory task (30 minute interval). Garro (1986) used Lucy and Shweder's equal discriminability array and found in three experiments on English speakers that the focal colours were remembered significantly better than the non-focal colours: the scores for the non-focal colours were similar to those found by Lucy and Shweder, but the scores for focal colours were about 20 per cent higher than those found by Lucy and Shweder. The results from the first three experiments were replicated on a small sample (10) of Tarascan–Spanish bilinguals from Mexico: the pattern of results was similar to the English speakers, but the scores for the Mexicans were about 15 per cent lower than for the English speakers. Garro interprets her results as consistent with the influences of perceptual salience and language.

The plausibility of colour perception being affected by language has been strengthened by psychophysical studies of perceptual learning (see Fahle, 1994, for a review). Sensitivity to simple perceptual attributes, such as vernier acuity (Fahle & Edelman, 1993), spatial frequency (Fiorentini & Berardi, 1981) and direction of motion (Ball & Sekuler, 1987), can be improved with practice, but such learning requires attention to the training stimuli—incidental retinal exposure is not sufficient (Shiu & Pashler, 1992). Further, learning to divide a perceptual dimension, such as brightness or saturation, into categories improves sensitivity to that dimension, as measured by same–different judgments (Goldstone, 1994). When an infant learns its mother tongue, it is likely that attention would be directed to the foci and to the boundaries of the basic colour categories encoded by the language. If sensitivity to colour is modifiable, this directed attention could result in greater sensitivity to the

foci and boundaries of the basic colour terms of the language than for other regions of colour space. This in turn implies that speakers of languages with different colour categories could have different distributions of sensitivity to colour across colour space in accord with the position of the significant category foci and boundaries within the respective languages. However, although it is possible that sensitivity to colour may be changeable, like other simple perceptual dimensions, there is no direct evidence yet that this is so.

Bornstein & Korda (1984, 1987) and Boynton, Fargo, Olson & Smallman (1989) have found categorical effects on hue judgments that parallel those found in judgments of acoustic similarity within and between phonemic categories. Using the 'same-different' judgments' paradigm, within-category 'different' responses are slower than comparable between-category 'different' responses. Thus, for instance, it takes longer to decide that two blue stimuli are different than to decide that a blue stimulus is different to a green stimulus, although the difference between the two pairs is matched in the number of Munsell hue steps (Bornstein & Korda, 1984, Expt 2).

The categorical effects found by Bornstein & Korda could be due to perceptual learning under the influence of language, as suggested above. Alternatively, such categorical effects could reflect inherited variations in sensitivity to different regions of colour space. If the categorical effect is due to language, then it should be less pronounced in speakers of a language that does not have a basic term category boundary within the region of colour space investigated than in speakers of a language that does have a basic colour term boundary within the region. Kay & Kempton (1984) compared the Tarahumara (from Mexico), whose language has a BLUE or GREEN term *siyóname*, with speakers of English on a 'triads' task. Participants were asked which of three colours from the BLUE-GREEN border region was the least similar to the other two. When two colours fell in either the English category *blue* or *green* and the third colour fell in the other category, English speakers were more likely than the Tarahumara speakers to select the third colour. Kay & Kempton interpret their results as due to language learning 'stretching' physical differences across category boundaries and thus supporting 'modest Whorfianism'. However, the number of participants was very small: there were four Tarahumara speakers and five English speakers. Further, there was no control task in which the same choices would be expected by the two groups on linguistic grounds.

We have carried out a partial replication of Heider's recognition memory task (Expt 3) on Setswana, English and Russian speakers (to be reported elsewhere). We used a much easier recognition memory task in which participants had to select a previously seen colour chip from either three or six alternatives. The stimuli were chosen so that in some cases differences were predicted on linguistic grounds, and in other cases no differences were expected on linguistic grounds. There were performance differences related to linguistic differences, but the most striking result was that the Setswana speakers' scores were very much lower than for the other two groups for all stimulus sets, and the Batswana group disliked the task. The performance differences could be due to linguistic effects on recognition memory—the Batswana finding the task difficult because Setswana does not provide sufficient available and effective labels—or it could be due to the strangeness of such

memory tasks. These results are similar to Heider's (1972) results with her Dani subjects; the Dani had much lower recognition scores than the American participants, but they also made their choices more quickly than the Americans. The pattern is consistent with a speed-error trade off, but it may also imply that the task was inappropriate for the Dani.

The reception our Batswana (the people of Botswana) informants gave to the recognition memory task, and the advice of our Batswana collaborators, led us to use an alternative task. We wanted a task that was culturally natural, and which was potentially sensitive to any influences of the colour lexicon. The task we chose was a simple colour grouping task: informants were presented with 65 colour 'tiles' that were a representative sample of 'colour space', and asked to sort the colours out into groups so that members of a group looked similar to each other. There were two versions of the task: the 'free sorting' task and a 'forced sorting' task. In the former (which we report here) we placed no constraint on how many groups were formed, whereas in the latter, the number of groups were specified (we report these data elsewhere). Pilot work showed that this was an acceptable and stress-free task. It has the potential to be sensitive to any linguistic effects, in that if learning to speak the language has changed colour perception, then this could be reflected in the relative similarity of the colours. And, if informants used linguistic labels to guide group formation, then speakers of languages with different colour categories might be expected to form different groups.

We compared speakers of English, Russian and Setswana using the basic colour grouping task to look for evidence of linguistic influences on colour perception or categorization. These languages vary in their number of basic colour terms, and there is also a particularly clear three-way difference in how they categorize the BLUE-GREEN region of colour space. English has 11 basic colour terms: *white, black, red, green, yellow, blue, brown, purple, pink, orange* and *grey* (Davies & Corbett, 1995). Russian has 12 basic colour terms (one more than Berlin & Kay's 1969 theoretical maximum): *belyj* 'white', *černyj* 'black', *krasnyj* 'red', *zelenyj* 'green', *želtyj* 'yellow', *sinij* 'dark blue', *goluboj* 'light blue', *koričnevyj* 'brown', *fioletovyj* 'purple', *rozovyi* 'pink', *oranževyj* 'orange', *seryj* 'grey' (Corbett & Morgan, 1988; Davies & Corbett, 1994; Morgan & Corbett, 1989). Setswana has five basic terms: *boswen* 'white', *bontsho* 'black', *bohhibidu* 'red', *botala* 'grue' and *borokwa* 'brown' (Davies *et al.*, 1992).

We used the colour grouping task to investigate two general questions. First, how similar would the grouping behaviour of the three language groups be? And second, would there be any evidence of systematic differences among the three language groups associated with language differences? As well as these general questions, we focus on two more specific ones. First, would the number of groups formed by the participants be related to the number of basic terms in the languages? Two lines of reasoning suggest that this might be so. If informants used a labelling strategy to form groups, it might be expected that the Batswana would form fewer groups on average than English or Russian speakers, as they have fewer basic colour terms to call on. And, if learning the language has amplified the perceptual distances between colours in different linguistic categories, this would increase the likelihood that colours from different categories would be placed in different groups, to maximize within group similarity. Second, each language differs from the other two in the

number of basic colour terms for the BLUE-GREEN region of colour-space: Setswana has a single term *botala* 'grue'; English has two basic terms, *blue* and *green*; and Russian has three basic terms—*sinij* 'dark blue', *goluboj* 'light blue' and *zelenyj* 'green'. If there are linguistic effects on colour grouping, then these should reflect the linguistic differences just described. Thus, the Batswana should be more likely to group BLUE with GREEN colours than speakers of either of the other two languages, and Russian speakers should be more likely than speakers of the other two languages to form groups distinguishing dark BLUES (*sinij* 'dark blue') from light BLUES (*goluboj* 'light blue').

Method

Participants

There were four samples of participants: an English-speaking sample from Britain; a Russian-speaking sample from the former Soviet Union; and two Setswana-speaking samples from Botswana in southern Africa—a main sample and a subsidiary sample. The samples are designated by the mother tongue of their members.

There were 47 people in the English sample, all of whom spoke English as their first language, although a few had various degrees of fluency in French. There were 24 men and 23 women, and their ages ranged from 21 to 65 years with a mean of 29 years. They lived in and around Guildford in the south of Britain. They were recruited from public places such as shopping centres. Most had left school when they were between 14 and 16 years of age and none had a university education. They were offered a small payment for their time, and any costs they incurred were refunded. There were 77 people in the Russian sample, all of whom spoke Russian as their first language and none were fluent in any other language. There were 24 men and 53 women and their ages ranged from 18 to 65 years with a mean of 34 years; they lived in Moscow. They were recruited from a number of districts in Moscow that varied in economic status. None had a university education. The participants were paid for their time. There were 44 people in the main Setswana sample and 40 people in the subsidiary sample; all informants were native Setswana speakers, and none knew more than a little English. They lived in small villages around the town of Kanye in the south of the country, and were members of the Bangwaketse tribe. Half were men, and half were women, and their ages ranged from 26 to 82 years, with a mean of 45 years. The demographic profiles of the two samples were very similar. The Setswana participants were recruited by first asking the local chief for permission to approach the headmen in local villages. The headmen were given a token gift, and they suggested who might be included in the sample. The participants were also given a gift (usually food) for participating. The majority of the Setswana sample had little schooling (none to three years at primary school in most cases).⁴

Stimuli

There were 65 coloured 'tiles' used in the grouping task. These were 50 mm squares of coloured paper mounted on thin (4 mm) plywood; the colours were sprayed with a light film of transparent varnish to protect them from staining during use. The colours were an evenly spread sample of 'colour space' taken from the Color-aid Corporation range. Appendix 1 gives their Color-aid codes, their CIE (Commission Internationale De Éclairage) tri-stimulus values (Y, x, y) under illuminant C (tables to convert CIE coordinates into Munsell notation are given by Newall, Nickerson & Judd, 1943, among others), and their CIE (1976) uniform chromaticity coordinates (L^*, u', v'), with the lightness values referenced to the white tile in the set. Figure 1*a* shows the location of the tile colours in CIE uniform

⁴ This is typical of most people in Botswana in this age range. Botswana now has a good primary and secondary school system available to all children, but this country-wide system is relatively new and would not have been available to most of our participants.

chromaticity space, and the loci of the universal foci (the achromatic foci have the same coordinates in the two-dimensional u', v' space, labelled grey in the figure, but differ on the lightness dimension).

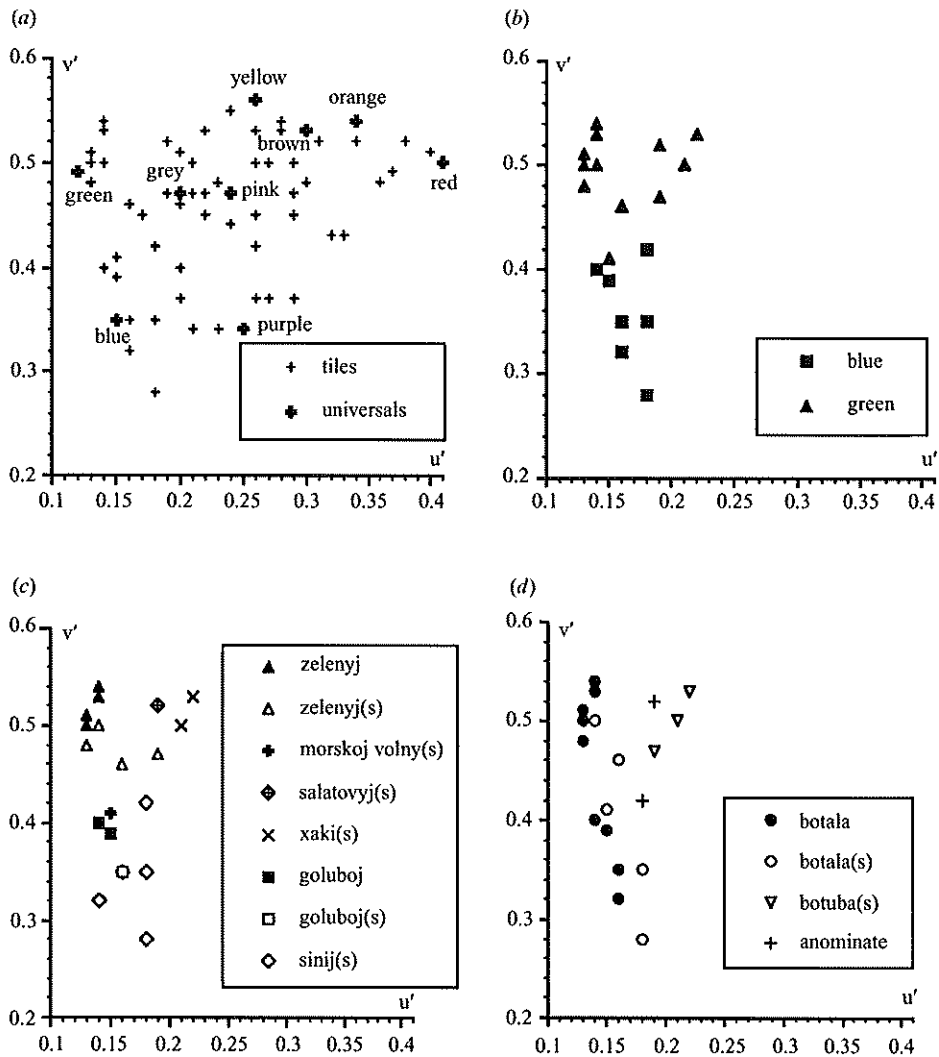


Figure 1. (a) Loci of the tile colours and the universal foci in CIE (1976) uniform chromaticity space. (b, c, d) Distribution of BLUE-GREEN terms in the CIE uniform chromaticity diagram for English, Russian and Setswana. (s) indicates a secondary term or an ancillary region of a basic term.

The Color-aid range is based on the Ostwald colour solid (see Foss, Nickerson and Granville, 1944, for an outline of this system) and is made up from 24 'hues': Y (yellow), O (orange), R (red), V (violet), B (blue) and G (green), and intermediate values designated by combinations of the previous codes; for instance YOY, YO, and OYO are the intermediate hues between Y and O. In addition to the hues, there are seven variants of each hue consisting of four 'tints', T1 to T4, and three 'shades', S1 to S3; the tints

have increasing amounts of white added to the hue as their index number increases, whereas the shades have increasing amounts of black added as their index number increases. In addition, there is a grey scale and several colours of particular significance to painters. The tiles were displayed on grey cloth whose CIE coordinates were $Y = 43.8 \text{ cdm}^{-2}$, $x = .39$, $y = .39$.

The City University Colour Vision Test (CUCVT; Fletcher, 1980) was used for assessing colour vision. This test requires no literacy, and pilot work had shown that the performance requirements of the test—matching to sample—were understood by respondents from each of the three countries.

Language and instructions

The experiment was conducted in the appropriate mother tongue for each sample. The English instructions were translated into each of the other languages by a native speaker of that language, and back-translated into English by another native speaker. The back-translation of Russian was satisfactory after the first cycle, but the Setswana version required some minor modifications before the back-translations were acceptable. The experiment was carried out by a native speaker of the appropriate language for the English and Setswana samples. The data from the Russian sample were collected by two experimenters: one was a native Russian speaker, and the other spoke fluent Russian, although his first language was English; the native Russian speaker tested 47 informants, and the remaining 30 informants were tested by the other experimenter.

Procedure

All participants in the main samples did five tasks: a colour term 'list task', the CUCVT, a 'free-sorting' task, a 'forced-sorting' task, and a colour naming task, in that order. The subsidiary Setswana sample only did a version of the colour naming task. The main focus of this paper is on performance on the free-sorting task, and its relationship to colour naming; data from the other tasks have been reported elsewhere (Davies *et al.*, 1992; Davies & Corbett, 1994, 1995; Davies, Laws, Corbett & Jerrett, submitted). Accordingly, here, we describe the free-sorting and colour naming tasks fully, but describe the remaining tasks in outline only. The first task—the list task—consisted of asking the participant to say as many colour terms as they knew; it typically took less than two minutes to complete. The CUCVT consists of 10 colour 'plates', each with a central colour spot and four equidistant surrounding colour spots. Respondents choose which of the four surrounding colour spots is most similar to the centre colour spot. The test usually takes less than two minutes. For the free-sorting task, the tiles were placed randomly, on a tray covered in the grey cloth, and participants were asked to sort the tiles into groups, so that 'members of a group looked similar to each other, just as members of a family resemble each other'. When the participants indicated that they were satisfied with their choices, the experimenter recorded which tiles were in each group, and then asked them each what their selections were based on. The time taken to do the task varied from about one minute to about seven minutes. Participants then sorted the tiles four more times, but on those occasions the number of groups they sorted the tiles into was specified by the experimenter. Finally, all of the tiles were arranged randomly on the grey cloth, and participants were asked to name as many of the tiles as they could; they picked up each tile they could name, and gave it to the experimenter, who recorded their response. They were encouraged to continue if they paused, but they were not forced to name all of the tiles.

The naming procedure resulted in the majority of the English and Russian participants naming most of the tiles, but on average the Setswana sample named a much smaller proportion of the tiles than the other two groups. Consequently, a further Setswana sample was used to produce more complete data on colour naming. This second sample was shown the full set of tiles in the same way as the previous group, and asked which of these tiles are x , where x was one of the Setswana colour terms offered frequently in the list task. This was done for each of the frequent terms in turn (see Davies *et al.*, 1992 for the full details). This version of the naming task produced a much higher response rate, although, as will be seen, the basic distribution of colour terms over the tile set did not differ between the two samples.

The circumstances under which the samples did the tasks varied somewhat. In Botswana, the task was done outside, avoiding direct sunlight or deep or dappled shade. The experimenter tried to ensure that the light levels were similar for each participant, but no formal measures of the illumination were taken.

In Moscow and in Britain, the experiment was carried out indoors. In Moscow, it took place in the participants' homes, in natural light, and again the experimenter tried to ensure that the light levels were similar across participants. In Britain, the experiment took place either in the participants' homes, or in their workplaces, or in the psychology department; in each case it took place under natural light, avoiding direct sunlight or deep or dappled shade. In all cases, although the light levels were not measured, they were sufficient for the experimenter to pass all items on the CUCVT easily.

Results

Colour naming

The main aim of this section is to establish how the BLUE-GREEN region is categorized in the three languages: this provides the basis for the tests of the influence of linguistic categorization on the free-sorting task, described in subsequent sections.

GREEN (*green*, or *zelenyj*), BLUE (*blue*, or *goluboj* or *sinij*) or a BLUE with GREEN (*botala* 'grue') terms were 'dominant' for 22 tiles for at least one of the samples (see Appendix 2). By dominant, we mean that half or more of the particular sample used the term to name the given tile.

Figures 1*b*, 1*c* and 1*d* illustrate the distribution of terms over the tiles in the CIE (1976) uniform chromaticity scale diagram (u' , v'). For English (Fig. 1*b*), all 22 of the tiles have a dominant term, and the space is sharply divided into just two categories—*green* and *blue*.

For Russian (Fig. 1*c*), there are three regions where tiles are labelled with dominant terms; the dominant terms are: *zelenyj* 'green', *goluboj* 'light blue' and *sinij* 'dark blue'. There are further regions where tiles have a 'most frequent term', but the frequency is not large enough for the term to be dominant. *Zelenyj* 'green', *goluboj* 'light blue' and *sinij* 'dark blue' each have such ancillary regions. In addition, the secondary terms shown in Fig. 1*c*—*morskoj volny* 'sea wave', *salatovyj* 'salad' and *xaki* 'khaki'—were all the most frequent term for the given tiles, and the term was used by between a quarter and a half of the sample. It can be seen that *morskoj volny* 'sea wave' lies between the *zelenyj* 'green' and the *goluboj* 'light blue' regions; in fact in the u' v' plane, BG-hue, for which *morskoj volny* 'sea wave' is the most frequent term, and GBG-S2, for which *goluboj* 'light blue' is the most frequent term, are superimposed. These tiles differ, however, on lightness (L^*) with GBG-S2 being lighter than BG-hue. *Salatovyj* 'salad' and *xaki* 'khaki' are secondary terms used to label tiles lying between the region for which *zelenyj* 'green' is dominant, and the region for which *zelyj* 'yellow' is dominant.

For Setswana, the only term used by more than a fifth of the sample to name a given tile, was *botala* 'grue'. In Fig. 1*d* we show four regions: first, the core *botala* 'grue' region, which is those tiles for which the term *botala* 'grue' is dominant; second, the ancillary *botala* 'grue' region, which is those tiles for which *botala* 'grue' was the most frequent term, and was used by at least a fifth of the sample; third, the region where *botuba* 'pale' was the most frequent term (albeit, with very low frequencies); and last, a region verging on the 'anominate', in which the highest frequency for any tile was two (which was for BV-S2 which two people named *botala* 'grue').

In summary, it can be seen that the BLUE-GREEN region is divided into three

dominant categories in Russian, into two dominant categories in English and into a single dominant category in Setswana. In addition the position of the category boundaries varies across languages. Even clear translation equivalents, such as *zelenyj* 'green' in Russian, and *green* in English, have boundaries in different positions (see Figs 1*b* and 1*c*). Further, from Figs 1*b* and 1*c*, it can be seen that, although the *blue* region overlaps with the joint *sinij-goluboj* 'dark blue–light blue' region, the position of the boundaries differ. And last, although *botala* 'grue' includes BLUE and GREEN (Fig. 1*d*) the position of its boundaries differs both from those of *blue* and *green*, and from *sinij* 'dark blue' and *goluboj* 'light blue' and *zelenyj* 'green'.

The free-sorting task

The number of groups. Table 1 summarizes the number of groups (N) that participants produced in the free-sorting task. It shows the mean and standard deviation across subjects for N , for each sample, and the number of basic terms in each language is included for comparison. It can be seen that the Setswana sample have the highest mean for N , and that the scores for English and Russian speakers are relatively similar to each other. One-way ANOVA indicated that there was a significant effect of language group on N ($F(2, 165) = 3.53, p < .05$). Further, the English and Russian scores do not differ significantly, but they are both significantly different to the Setswana scores (protected $t(165) > 2.1, p < .05$). The English and Russian samples have mean scores which are similar to the number of basic terms in the respective languages, but the mean for the Setswana sample is about three times the number of basic terms in Setswana.

Tile similarity matrices: Global structure. Tile similarity matrices were constructed for each sample. The cell entries in the matrices are the percentage of participants in each sample that placed a given pair of tiles in the same group, for each possible pair of tiles ($65 \times 64/2$ pairs). Thus the similarity scores could range from 0 per cent to 100 per cent, with a high score indicating that a large proportion of participants placed a given pair of tiles in the same group, and a low score indicating that few had placed that pair of tiles in the same group. The correlations between the three matrices ranged from .83 for English–Setswana, to .95 for English–Russian, indicating that the structure of the matrices were very similar across languages, despite the differences between linguistic categories.

Consensus in grouping. Although the basic structure of the three matrices is similar, they could still vary in the degree of consensus across the informants in each sample. The standard deviation of the similarity matrix (SD_m) is an index of consensus across participants over which tiles were grouped together: the greater the consensus, the closer the score would be to either 0 per cent (participants never grouped that pair of tiles together) or to 100 per cent (participants always grouped that pair of tiles together); the maximum possible SD_m is, for the case of perfect consensus, where all scores are either 0 per cent or 100 per cent. The values for SD_m were: 6.5 for Setswana, 10.9 for English and 14.2 for Russian. Each of these scores is significantly different to the others at at least the 1 per cent level (minimum $F(20, 79) = 1.70$).

Table 1. Mean and standard deviation (SD) of the number of groups produced in the free-sorting task for each sample and the number of basic colour terms in each language

	English	Russian	Setswana
Mean	12.4	12.9	15.9
SD	6.9	7.3	5.9
Number of basic terms	11	12	5

However, SD_m may be influenced by N , and by the distribution of the number of tiles in the N groups, but the relative differences in consensus remain even allowing for differences in N .⁵

Equality of group sizes. The three languages also differ in how equal the group sizes are (the distributions of the number of tiles in a group across the N groups). The mean values of SD_g were 1.96, 2.96 and 3.53, for Setswana, Russian and English respectively; ANOVA indicated that there was a significant effect of group $F(2, 163) = 4.45, p < .0001$). Both Russian and English are significantly different to Setswana, but not to each other ($p < .05$ level, Tukey HSD *post hoc* test).⁶

Multidimensional scaling. The structure of the similarity matrices was investigated by non-metric multidimensional scaling (MDS). In each case three-dimensional solutions yielded good fits to the original data (Kruskall's stress values were between 0.12 and 0.17, and the r^2 's between the derived and the original distances were between 0.81 and 0.91). Figures 2 to 4 show the MDS solutions for the English, Russian and Setswana speaking samples for the first two dimensions; Colour-aid codes are used to designate the tiles. (In order to save space the label hue has been omitted; thus,

⁵ If the tiles are distributed across the N groups as equally as possible, then the maximum possible value of SD_m decreases with N . Thus it could be that the smaller SD_m for Setswana is due to Setswana having the largest value for N . On the other hand, if the distribution of the number of tiles in a group is maximally skewed ($N-1$) groups of one tile, and one group of $65-(N-1)$ tiles, then the maximum possible value of SD_m increases with N . We estimated the maximum possible value of SD_m for the three languages, using N , and the distribution of the number of tiles in a group, in the following way. For each participant, we calculated the ratio of the standard deviation of the number of tiles in a group (SD_g) to the maximum possible SD_g for the observed value of N . If the tiles were distributed as equally as possible, then SD_g would tend to zero; on the other hand, the maximum possible SD_g would occur for the case of maximum skew. Thus, the ratio is an index of how skewed the distributions were. We then used the mean ratio across participants, for each language, to interpolate between the maximum possible SD_m for equal group sizes and the maximum possible SD_m for skewed distributions, to give us an estimate of the maximum possible SD_m for each language. Finally, we calculated the ratio of the observed SD_m to the maximum possible SD_m . The ratios were: .26 for Setswana; .36 for English; and .50 for Russian—the same rank order as for the observed SD_m 's. Thus the consensus measure incorporating influences due to differences in N and SD_g is about 40 per cent greater than for English than Setswana, while the Russian score is about 90 per cent greater than the Setswana score, and about 40 per cent greater than the English score.

⁶ The scores for the ratio of SD_g to the maximum possible SD_g also differed: the mean ratios were 0.15, 0.18 and 0.22 for Setswana, Russian and English respectively; these scores are significantly different ($F(2, 163) = 22.94, p < .0001$). All three groups are significantly different to each other ($p < .05$ level; Tukey HSD).

for instance, in Fig. 2, V and VRV which can be found to the left of the plot with positive *y* values, stand for V-hue and VRV-hue respectively.

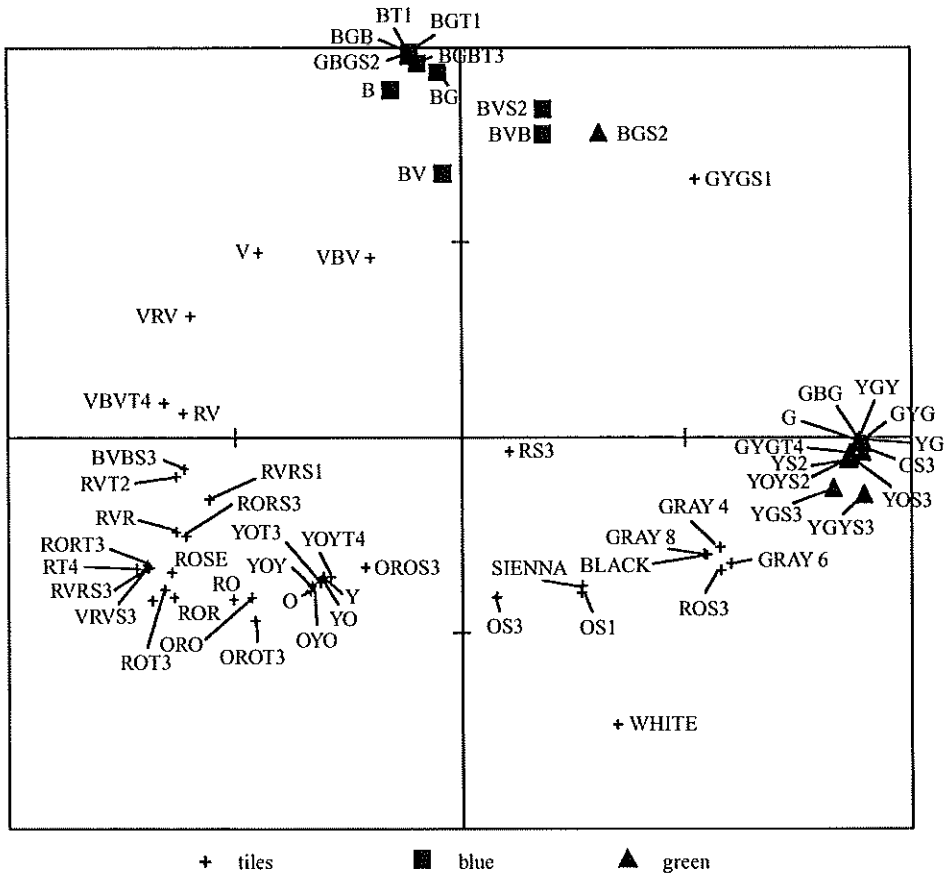


Figure 2. English MDS solutions (see Fig. 1 for location of terms in CIE coordinates).

The basic structure of the three solutions is similar, as would be expected given the high correlations between the matrix scores that we reported above. In each case, the spatial distribution of colours resembles the classic colour circle: REDS (e.g. R, RO etc.) are opposite GREENS (e.g. G, GBG etc.) and BLUES (e.g. B, BVB etc.) are opposite YELLOWS (e.g. Y, YO etc.); and hues are ordered around the circumferences of the 'circles' in the sequence RED, PURPLE, BLUE, GREEN, YELLOW, ORANGE.

Inspection of the loci of the BLUE and GREEN tiles (tile codes are in Appendix 1; see also Fig. 1) reveals differences between the BLUE-GREEN regions in the English and Russian plots (Figs 2 and 3) and the Setswana plot (Fig. 4). The BLUE-GREEN region in the English plot is divided into two major clusters, one with *green* as the most frequent term, and one with *blue* as the most frequent term. The pattern of the data from the Russian sample resembles that from the English sample, provided we

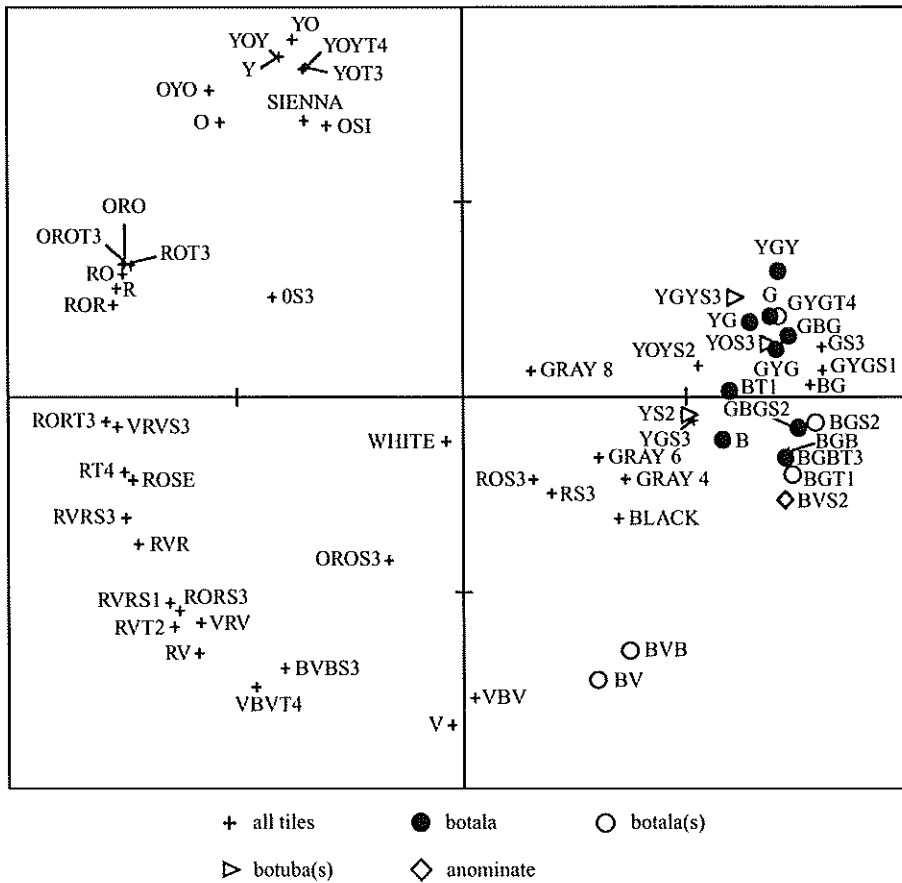


Figure 4. Setswana MDS solutions (see Fig. 1 for location of terms in CIE coordinates).

Russian sample, even though all tiles in both sets have *blue* as their dominant term in English.

Category boundary effects. The MDS solutions suggest that there are differences in how the BLUE–GREEN tiles are grouped in the three languages. In order to investigate whether these impressions are statistically reliable, we next describe a series of ANOVAs carried out on various sections of the similarity matrices for just the BLUE–GREEN tiles: the key planned comparisons, outlined in the introduction, are between, first, the intra-GREEN and the intra-BLUE regions and the inter-BLUE–GREEN region; and second, for the BLUE region (between the intra-DARK BLUE and intra-LIGHT BLUE regions, and the inter-DARK BLUE–LIGHT BLUE region). In addition we also look for possible effects around the YELLOW–GREEN boundary.

The main statistical analysis was based on two-way ANOVAs on colour region by language, with repeated measures on the first factor. The crucial test is whether there is a significant interaction between the two factors and if there is, whether the pattern

of the interaction is consistent with the predictions. The main effects are highly significant in each analysis, and thus in order to save space, we concentrate on reporting the interaction. We interpret the interactions using simple main effects or *post hoc* tests (Tukey's HSD); all differences we describe are significant at 5 per cent level at least.

Partitioning of BLUE-GREEN. Table 2A shows the mean scores for the three core regions—BLUE, GREEN and BLUE by GREEN—for the three languages. (The cores are those tiles for which either a BLUE or a GREEN or a GRUE term is dominant.)

Table 2. Mean similarity scores for three border regions for each language (standard deviations are given in parentheses)

Colour region	Language group			Mean
	English	Russian	Setswana	
A				
Intra-BLUE	85.8 (25.1)	74.0 (32.3)	64.7 (32.2)	74.8 (32.2)
Intra-GREEN	96.2 (11.2)	90.1 (23.9)	75.2 (30.8)	87.2 (24.6)
BLUE by GREEN	0.0 (0.0)	2.5 (2.0)	7.4 (3.7)	3.3 (14.5)
Mean	60.7	55.5	49.1	55.1
B				
Intra-DARK BLUE	78.3 (31.1)	73.1 (32.6)	41.6 (27.7)	64.3
Intra-LIGHT BLUE	89.7 (21.8)	80.0 (29.8)	65.5 (33.4)	78.4
DARK BLUE by LIGHT BLUE	71.1 (37.1)	58.6 (38.2)	37.5 (27.7)	55.7
Mean	79.7	70.6	48.2	66.1
C				
YELLOW-GREEN by BLUE	0.0 (0.0)	2.5 (9.4)	2.7 (8.4)	1.7
YELLOW-GREEN by GREEN	71.3 (36.9)	39.7 (39.3)	11.4 (12.1)	40.8
Intra-YELLOW-GREEN	68.8 (31.2)	55.2 (34.8)	34.2 (20.8)	52.7
Mean	46.7	32.5	16.1	31.7

As can be seen from Table 2A, the mean BLUE by GREEN similarity is very much less than either the mean BLUE by BLUE similarity or the mean GREEN by GREEN similarity, and overall, the mean similarity for Setswana is less than for Russian, which in turn is less than for English. In addition, the interaction between colour

region and language is significant ($F(4, 326) = 7.0, p < .0009$). Setswana has the highest score for the BLUE by GREEN region but the lowest scores for the GREEN and the BLUE regions.

Partitioning of BLUE. Table 2B shows the mean scores for each language for each of the three submatrices intra-DARK BLUE (BV-hue to BGB-hue); intra-LIGHT BLUE (B-T1 to GBG-S2) and DARK BLUE by LIGHT BLUE. In English, all of these tiles have *blue* as their dominant term, whereas in Russian, the DARK BLUE tiles fall into the *sinij* category, and the LIGHT BLUE tiles fall into the *goluboj* category. In Setswana, all of the tiles except BV-hue and BVB-hue, fall into the *botala* category: thus the *botala* category boundary falls in the DARK BLUE region. Thus if there are linguistic effects on grouping, both Russian and Setswana should be less likely to group DARK BLUE tiles with LIGHT BLUE tiles than English.

The data in Table 2B seem consistent with this prediction. However, the ANOVA indicates that although both main effects are highly significant, the interaction between colour region and language is not ($F(4, 326) = 2.2, p = .07$).

The GREEN-YELLOW boundary. The four tiles, YGY-S3, YG-S3, Y-S2 and YOY-S2, lie close to the GREEN boundary, towards YELLOW. In English, these four tiles fall within the *green* category, whereas in Russian, there are three different most frequent terms used in the region: YGY-S3 has *salatovyj* 'salad' as the dominant term; YG-S3 has *zelenyj* 'green' as the most frequent term; and Y-S2 and YOY-S2 have *xaki* 'khaki' as the dominant and the most frequent terms respectively. Lastly, in Setswana, these tiles have no dominant term (but they are definitely not in the *botala* 'grue' category). Table 2C shows the mean similarities between this YELLOW-GREEN region and the core BLUE region (B-hue to BG-hue) and between the YELLOW-GREEN region and the core GREEN region (GBG-hue to YGY-hue), together with the intra-YELLOW-GREEN similarity.

The English sample (for whom all the tiles are *green*) never group a YELLOW-GREEN tile with a BLUE tile, whereas they group the YELLOW-GREEN tiles with GREEN tiles, and with each other, about 70 per cent of the time. The Setswana group (for whom the BLUE tiles and the GREEN tiles are both *botala* 'grue') like the English group, rarely group the YELLOW-GREEN tiles with the BLUE tiles, but unlike the English group, they also tend not to group the YELLOW-GREEN tiles with the GREEN tiles, although they do so more often than with the BLUE tiles. Further, although the Setswana score for the YELLOW-GREEN by YELLOW-GREEN set is the highest of the Setswana scores, it is lower than the equivalent scores for either of the other two groups. The Setswana data suggests that tiles that fall outside the *botala* 'grue' boundary tend not to be grouped with tiles that are included in *botala* 'grue'. The Russian data shows a pattern that falls between the other two languages. Russian tends not to group the YELLOW-GREEN tiles with the BLUE tiles, but they do so about 40 per cent of the time with the GREEN tiles, even though the tiles do not fall strongly into the *zelenyj* 'green' category. The Russians do tend to group the YELLOW-GREEN tiles together (mean score = 55.2 per cent), but some of the tiles fall into common secondary term categories for the Russian group.

The foregoing impressions are supported by the ANOVA. The interaction between

colour region and language is highly significant ($F(4, 326) = 22.0, p < .0009$). Tests of the simple main effects of language at each level of colour region show that there is no difference between languages for the YELLOW-GREEN by BLUE region. In contrast, the other two simple main effects are highly significant ($p < .00009$) and *post hoc* tests indicate that each pairwise difference between languages is significant.

Discussion

The overall pattern of results shows that the behaviour of the three language samples is broadly similar. The correlations between the similarity matrices were all high (.83 to .95), and correspondingly the broad structure of the MDS solutions were similar to each other. However, as well as the large-scale similarity across languages, there are also consistent differences. There are 'structural' differences in how the tiles are grouped (the number of groups produced, the equality of group sizes, and the levels of consensus over grouping), as well as 'category effects'. The structural differences were not predicted, and they do not seem to be related directly to the category effects. We first discuss the structural effects and then consider the category effects.

There were three structural differences among the samples in how the tiles were grouped. First, Setswana, counter to our prediction, produced more groups on average than English or Russian. Setswana produced about 16 groups compared to about 12 for English and 13 for Russian. Thus, the Setswana data does not support the conjecture that the number of groups produced would reflect the number of basic colour terms in the language: 16 is about three times the number of basic colour terms in Setswana. On the other hand, the mean number of groups formed for English and Russian are close to the number of basic colour terms in the respective languages (11 for English and 12 for Russian). Second, there were reliable differences between languages in the level of agreement over which tiles to group together: Setswana showed the least consensus and Russian the most. Third, Setswana tended to produce more equal group sizes than the other languages, and there is some evidence that English produced less equal group sizes than Russian.

Turning to the category effects, the MDS solutions (Figs 2-4) show that BLUE and GREEN colours are more likely to be grouped together in Setswana than in either English or Russian. This was supported by the analysis of how often BLUE tiles were grouped with GREEN tiles (Table 2A): the differences were small, but statistically significant. None of the English sample grouped a BLUE tile with a GREEN tile compared to about 5 per cent of the Russian sample, and about 20 per cent of the Setswana sample. On the other hand, there was no support for the prediction that Russian would be less likely to group DARK BLUE tiles with LIGHT BLUE tiles than the other language samples, either from the MDS (Figs 2-4) or from the detailed analysis of the intra-BLUE submatrix (Table 2B).

There is also support for the effect of language on grouping to be found in the detailed exploration of boundary positions within the BLUE-GREEN region. Setswana was less likely than English or Russian to group YELLOW-GREEN tiles with GREEN tiles, and Russian was less likely to do so than English (Table 2C). In Setswana, the YELLOW-GREEN tiles fall outside the *botala* 'grue' category, whereas the GREEN tiles are included in *botala* 'grue'. In Russian, the YELLOW-GREEN region is shared between

three colour terms: *salatovij* 'salad', *xaki* 'khaki' and *zelenyj* 'green'. In English, the YELLOW-GREEN region is included in *green*. Thus the differences in grouping across languages can be described as Setswana not grouping non-*botala* tiles with *botala* tiles, and Russian not grouping non-*zelenyj* tiles with *zelenyj* tiles. We did not predict this effect, but in retrospect we should have. We predicted that Setswana would tend to group BLUE with GREEN because they are both included in *botala*, and by implication that English and Russian would tend to group GREEN tiles together because they were included in either *green* or *zelenyj*. The complement of this prediction is that non-*botala* tiles should not be grouped with *botala* tiles, and that non-*zelenyj* tiles should not be grouped with *zelenyj* tiles. In fact, the evidence in support of the complementary prediction is clearer than the evidence in favour of the original prediction (compare Table 2A with Table 2C).

The data in Table 2B can be interpreted in a similar manner to that derived from the complementary hypothesis suggested above. Two of the CORE DARK BLUE tiles (BV-hue and BV-S2) have *botala* 'grue' as their most frequent term, but with scores of just 10 and 2. Thus, not grouping 'weak' *botala* tiles with *botala* tiles may be why intra-DARK BLUE and DARK BLUE by LIGHT BLUE scores are lower for Setswana than for the other languages.

We believe that we have established that there are reliable differences between the samples in how the tiles are grouped. The size of these effects is often small relative to the overall similarity between language groups, but they are associated with differences in the colour category structures between the languages. The data supports the modest linguistic relativity of Kay & Kempton (1984). However, it is unclear what the locus (or loci) of the effect is. Perhaps the least interesting possibility is that respondents use a direct language strategy to group the tiles. If this was so, then language would be affecting behaviour, but strictly speaking it may not be affecting cognition. A more interesting possibility is that learning the language has in some sense changed some aspect of colour perception. When asked on what basis did they perform the task, our respondents almost always said they used a strategy based on colour appearances, and nobody confessed to using a language strategy. However, it is important to seek corroborating evidence from other colour cognition tasks before concluding that language may affect colour perception. We are carrying out a series of experiments that vary the relative perceptual and memory loads in order to locate the origin of the effect more directly.

Can what we have called structural differences be explained in terms of linguistic differences, or must we look elsewhere for an explanation? There are three structural effects: the number of groups, the distribution of number of tiles in a group, and the consensus across respondents over which tiles to group together. These differences between English and Russian on the one hand, and Setswana on the other, could be attributed to the relatively low salience of the notion of colour in Setswana, relative to the other two languages. Setswana has fewer basic colour terms than English or Russian; there is less agreement over what colour terms denote than in English or Russian (Davies *et al.*, 1992; Davies & Corbett, 1994, 1995); the age at which children master the available colour terms is greater than in English or Russian (Davies *et al.*, 1994); and the general significance of the concept of colour is less than for English or Russian (Davies *et al.*, 1992). It may be that the lower consensus level

in the grouping task for Setswana simply reflects the lower availability of a shared framework (possibly a linguistic framework) to constrain how the tiles are grouped. On the other hand, such a salience explanation does not account for the lower consensus levels of English relative to Russian.

The tendency of Setswana to produce more groups than English or Russian may also be attributed to salience effects. Perhaps without a salient categorical framework, the manifest perceptual differences between the tiles are attended to at the expense of attending to their categorical attributes. The average inter-tile similarity within a group tends to decrease as the number of tiles in a group increases. Thus, a grouping strategy which aimed to maximize the average within-group perceptual similarity would tend to force the number of groups up towards 65. The availability of a salient category structure might serve to dampen this tendency by allowing categorical equivalence to offset perceptual differences.

Variations in the availability of a salient category structure might also account for the differences in the distribution of group sizes. If there were no category structure available and there were no 'perceptual boundaries' in the colour space occupied by the 65 tiles, then it might be expected that respondents would tend to form equal sized groups, as measured by both the number of tiles and the volume of colour space. (The CIE $L^* u' v'$ space is intended to be an approximately perceptually equal space, although this ignores the possibility of individual or cultural differences; distances in the space represent perceptual distances. The tiles are evenly spread in $L^* u' v'$ and thus equal volumes would also result in equal numbers in the groups.) Basic colour categories do not segment colour space into equal sized categories. For instance, the RED, WHITE and BLACK categories occupy smaller regions than do the BLUE or GREEN or GRUE categories (Davies *et al.*, 1992; Davies & Corbett 1994, 1995). Thus, to the extent that grouping was performed categorically (groups consisting of tiles with shared names), the variability in group sizes would tend to increase. It may be that the greater salience of colour in English and Russian relative to Setswana makes it more likely that a categorical strategy is used. The effect of the availability of such a categorical structure on the variability of group sizes may add to the effects due to any perceptual 'distortions' of the homogeneity of the colour space.

It is common in cross-cultural studies to compare only two languages (e.g. Heider, 1972; Kay & Kempton, 1984). Had we only studied English and Setswana, the data would have supported arguments for linguistic effects on colour cognition relatively clearly. As it is, the inclusion of Russian—a language similar to English as measured by the number of basic colour terms—makes the patterns in the data considerably more complex. This may suggest that caution is required in interpreting comparisons based only on two languages. The differences between English and Russian may indicate that there are important influences on colour grouping in addition to the categorical influences and the general salience effects we have discussed.

It is possible that the reason(s) for the differences in colour grouping between the samples may not be linguistic. It is impossible to match samples on all but the independent variable in the kind of cross-cultural study we have reported here. Our three samples differed in their level of education, the physical environments in which they lived, their domestic arrangements, the prevailing climate and so on. There may also be prevailing cultural norms or habits that account for the structural differences

we found. We have further evidence that the main colour grouping effect—the Batswana tendency to group BLUE and GREEN—is reliable and general. In the forced sorting task (the second phase of the current experiment) the Batswana were again much more likely to group BLUE and GREEN than either the English or Russians. Further, in a triads task (cf. Kay & Kempton, 1984) the differences we found between English and Setswana speakers were restricted to regions of colour space where there were differences in the positions of category boundaries in the two languages. This suggests that it is unlikely to be some general cultural difference in the nature of choice or categorization producing the results. On the other hand, we are in a less strong position to defend our conjecture that the structural differences found arise from linguistic differences. There could be general cultural or educational difference responsible for these effects.

Even assuming that the colour grouping differences are reliably associated with linguistic differences does not preclude the possibility that cultural or environmental differences gave rise to the linguistic differences in the first place. Presumably there has been less to gain in Botswana from encoding the difference between BLUE and GREEN in the language than there has been in the UK or Russia. Whatever the reasons, they may still be present in the culture in general, reinforcing the specific cultural effect of language.

In summary, colour grouping is very similar in the three languages, supporting perceptual universalism. On the other hand, there are also small but reliable differences in colour grouping that are consistent with linguistic differences influencing grouping. In addition there are structural differences in grouping among the three languages, which may reflect the differential availability of colour category structures in the three languages. To a first approximation colour perception is universal, but there may be scope for small-scale modulations by language and other cultural influences.

Acknowledgements

This work was supported by the ESRC (grant R000 23 1958) which we gratefully acknowledge. We are indebted to our collaborators in Botswana—David and Tiny Jerrett—and in Russia—Vladimir Moss and Svetlana Stepanskaja. Catriona MacDermid helped to organize the data collection and with the initial data analysis, and we are grateful for her contribution. We would also like to thank two anonymous reviewers for helpful comments on the originally submitted version of this paper.

References

- Ball, K. & Sekuler, R. (1987). Direction-specific improvement in motion discrimination. *Vision Research*, 27(6), 953–965.
- Berlin, B. & Kay, P. (1969). *Basic Color Terms: Their Universality and Evolution*. Berkeley and Los Angeles, CA: University of California Press.
- Bornstein, M. H. (1973). Color vision and color naming: A psychophysiological hypothesis of cultural difference. *Psychological Bulletin*, 80, 257–285.
- Bornstein, M. H. & Korda, N. O. (1984). Discrimination and matching within and between hues measured by reaction times: Some implications for categorical perception and levels of information processing. *Psychological Research*, 46, 207–222.
- Bornstein, M. H. & Korda, N. O. (1987). Some psychological parallels between categorisation processes in vision and audition. In S. Harnad (Ed.), *Categorical Perception*. New York: Cambridge University Press.

- Boynton, R. M., Fargo, L., Olson, C. X. & Smallman, H. S. (1989). Category effects in color memory. *Color Research and Application*, 14, 229–234.
- Brown, R. (1976). In memorial tribute to Eric Lenneberg. *Cognition*, 4, 125–153.
- Brown, R. & Lenneberg, E. H. (1954). A study in language and cognition. *Journal of Abnormal and Social Psychology*, 49, 454–462.
- Corbett, G. G. & Morgan, G. (1988). Colour terms in Russian: Reflections of typological constraints in a single language. *Journal of Linguistics*, 24, 31–64.
- Davies, I. R. L. & Corbett, G. G. (1994). The basic color terms of Russian. *Linguistics*, 32, 65–89.
- Davies, I. R. L. & Corbett, G. G. (1995). A practical field method for determining basic colour terms. *The World's Languages*, 9, 25–36.
- Davies, I. R. L., Laws, G., Jerrett, D. & Corbett, G. G. (submitted). Cross-cultural differences in colour vision: Acquired colour blindness in Africa.
- 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.
- De Valois, R. L. & De Valois, K. K. (1993). A multi-stage color model. *Vision Research*, 33, 1053–1065.
- Fahle, M. (1994). Human pattern recognition: Parallel processing and perceptual learning. *Perception*, 23, 411–427.
- Fahle, M. & Edelman, S. (1993). Long-term learning in vernier acuity: Effects of stimulus orientation, range and of feedback. *Vision Research*, 33, 397–412.
- Fiorentini, A. & Berardi, N. (1981). Learning in grating waveform discrimination: Specificity for orientation and spatial frequency. *Vision Research*, 21, 1149–1158.
- Fletcher, R. (1980). *The City University Colour Vision Test*, 2nd ed. London: Keeler.
- Fodor, J. A. & Pylyshyn (1981). How direct is visual perception?: Some reflections on Gibson's 'Ecological Approach'. *Cognition*, 9, 139–196.
- Foss, C. E., Nickerson, D. & Granville, W. C. (1944). Analysis of the Ostwald Colour System. *Journal of the Optical Society of America*, 34(7), 361–381.
- Garro, L. C. (1986). Language, memory, and focality: A reexamination. *American Anthropologist*, 88, 128–136.
- Gleason, H. A. (1961). *An Introduction to Descriptive Linguistics*. New York: Holt Rinehart & Winston.
- Goldstone, R. (1994). Influences of categorisation on perceptual discrimination. *Journal of Experimental Psychology: General*, 123(2), 178–200.
- Heider, E. R. (1972). Universals in color naming and memory. *Journal of Experimental Psychology*, 93(1), 10–20.
- Hering, E. (1920). *Outlines of a Theory of the Light Sense*. Translated from German by L. M. Hurvich & D. Jameson (1964). Cambridge: Harvard University Press.
- Jameson, D. (1985). Opponent-colours theory in the light of physiological findings. In D. Ottoson & S. Zeki (Eds), *Central and Peripheral Mechanisms of Colour Vision*. London: Macmillan.
- Kay, P. & Kempton, W. (1984). What is the Sapir-Whorf hypothesis? *American Anthropologist*, 86, 65–79.
- Kay, P. & McDaniel, C. K. (1978). The linguistic significance of the meanings of basic color terms. *Language*, 54, 610–646.
- Koffka, K. (1935). *Principles of Gestalt Psychology*. New York: Harcourt Brace.
- Lantz, D. & Steffire, V. (1964). Language and cognition revisited. *Journal of Abnormal and Social Psychology*, 69, 472–481.
- Levinson, S. C. & Brown, P. (1994). Immanuel Kant among the Tenejapans: Anthropology as empirical philosophy. *Ethos*, 22(1), 3–41.
- Lucy, J. A. (1992). *Language Diversity and Thought: A Reformulation of the Linguistic Relativity Hypothesis*. *Studies in the Social and Cultural Foundation of Language No 12*. Cambridge: Cambridge University Press.
- Lucy, J. A. & Shweder, R. A. (1979). Whorf and his critics: Linguistic and nonlinguistic influences on color memory. *American Anthropologist*, 81, 581–615.
- Morgan, G. & Corbett, G. (1989). Russian colour term salience. *Russian Linguistics*, 13, 125–14.
- Newhall, S. M., Nickerson, D. & Judd, D. B. (1943). Final report of the OSA Subcommittee on the spacing of the Munsell colors. *Journal of the Optical Society of America*, 33, 385–418.
- Ratner, C. (1989). A sociohistorical critique of naturalist theories of color perception. *Journal of Mind and Behaviour*, 10, 361–372.

- Rivers, W. H. R. (1901). Introduction. In A. C. Haddon (Ed.), *Reports on the Cambridge Anthropological Expedition to the Torres Straits*. Cambridge: Cambridge University Press.
- Shiu, L. & Pashler, H. (1992). Improvement in line orientation is retinally local but dependent on cognitive set. *Perception and Psychophysics*, **52**, 582-588.
- Simpson, C. (1991). Colour perception: Cross-cultural linguistic translation and relativism. *Journal for the Theory of Social Behaviour*, **21**, 409-429.
- Turton, D. (1980). There's no such beast: Cattle and colour naming among the Mursi. *Man*, **15**, 320-338.
- Turvey, M. T. (1977). Contrasting orientations to the theory of visual information processing. *Psychological Review*, **84**(1), 67-88.
- Whorf, B. (1956). *Language, Thought and Reality*. Cambridge, MA: MIT Press.

Received 9 November 1995; revised version received 8 May 1996

Appendix 1

Color-aid codes and CIE coordinates for the tile colours

Color-aid code		CIE coordinates					
		Y	x	y	L*	u'	v'
Y	hue	64.77	0.47	0.48	91.49	0.24	0.55
	S2	16.99	0.41	0.44	52.81	0.22	0.53
YOY	hue	47.48	0.50	0.43	80.92	0.28	0.54
	T4	55.63	0.45	0.41	86.18	0.26	0.53
YO	S2	22.08	0.36	0.38	59.09	0.21	0.50
	hue	39.52	0.51	0.41	75.17	0.30	0.53
	T3	47.02	0.48	0.41	80.61	0.28	0.53
OYO	S3	10.72	0.36	0.41	43.02	0.20	0.51
	hue	26.51	0.54	0.37	63.81	0.34	0.52
O	hue	25.00	0.54	0.37	62.26	0.34	0.52
	S1	14.34	0.50	0.37	49.03	0.31	0.52
ORO	S3	9.15	0.42	0.36	39.98	0.26	0.50
	hue	18.87	0.57	0.34	55.26	0.38	0.52
	T3	36.88	0.46	0.35	73.09	0.29	0.50
RO	S3	26.51	0.33	0.32	63.81	0.21	0.47
	hue	16.22	0.58	0.33	51.75	0.40	0.51
	T3	32.66	0.45	0.32	69.56	0.30	0.48
ROR	S3	4.19	0.37	0.34	27.15	0.23	0.48
	hue	15.23	0.53	0.31	50.35	0.37	0.49
	T3	29.82	0.42	0.30	67.00	0.29	0.47
R	S3	20.71	0.34	0.28	57.50	0.24	0.44
	hue	11.71	0.50	0.29	44.78	0.36	0.48
	T4	24.34	0.40	0.27	61.57	0.29	0.45
RVR	S3	4.81	0.33	0.30	29.18	0.22	0.45
	hue	9.11	0.42	0.24	39.90	0.33	0.43
	S1	12.79	0.35	0.25	46.60	0.26	0.42
RV	S3	28.43	0.36	0.28	65.69	0.26	0.45
	hue	6.97	0.33	0.19	35.13	0.29	0.37
VRV	T2	14.51	0.31	0.19	49.28	0.27	0.37
	hue	6.71	0.30	0.19	34.48	0.26	0.37

	S3	8.42	0.36	0.28	65.68	0.26	0.45
V	hue	4.67	0.26	0.17	28.74	0.23	0.34
VBV	hue	4.13	0.24	0.17	26.94	0.21	0.34
	T4	19.05	0.25	0.20	55.49	0.20	0.37
BV	hue	4.21	0.22	0.19	27.22	0.18	0.35
	S2	7.88	0.25	0.26	37.26	0.18	0.42
BVB	hue	4.80	0.19	0.13	29.15	0.18	0.28
	S3	26.65	0.26	0.23	63.95	0.20	0.40
B	hue	9.51	0.18	0.16	40.71	0.16	0.32
	T1	19.02	0.20	0.19	55.45	0.16	0.35
BGB	hue	9.62	0.19	0.19	40.93	0.16	0.35
	T3	23.08	0.20	0.23	60.21	0.15	0.39
BG	hue	8.93	0.20	0.25	39.53	0.14	0.40
	T1	16.57	0.19	0.25	52.24	0.14	0.40
	S2	7.42	0.21	0.26	36.21	0.15	0.41
GBG	hue	10.69	0.23	0.37	42.96	0.13	0.48
	S2	20.79	0.20	0.25	57.60	0.14	0.40
G	hue	11.99	0.24	0.42	45.26	0.13	0.50
	S3	6.10	0.26	0.33	32.91	0.16	0.46
GYG	hue	12.89	0.25	0.44	46.76	0.13	0.51
	T4	31.14	0.26	0.41	68.21	0.14	0.50
	S1	15.59	0.26	0.31	50.86	0.17	0.45
YG	hue	14.66	0.28	0.48	49.51	0.14	0.53
	S3	5.78	0.30	0.34	32.04	0.19	0.47
YGY	hue	18.92	0.30	0.51	55.32	0.14	0.54
	S3	35.87	0.35	0.43	72.27	0.19	0.52
Rose red		17.63	0.41	0.24	53.66	0.32	0.43
Sienna		13.31	0.44	0.36	47.43	0.27	0.50
White		81.40	0.32	0.33	100.00	0.20	0.47
Gray 1		47.55	0.32	0.33	80.97	0.20	0.47
Gray 2		30.59	0.32	0.33	67.71	0.20	0.47
Gray 4		18.88	0.31	0.31	55.27	0.20	0.46
Gray 6		11.20	0.31	0.31	43.89	0.20	0.46
Gray 8		4.53	0.31	0.32	28.89	0.20	0.46
Black		3.59	0.34	0.33	24.98	0.22	0.47

Appendix 2

Most frequent terms, and the percentage of the sample that used them, for tiles in the BLUE-GREEN region, for the three languages (for Setswana, data is for the subsidiary sample)

Color-aid code	English		Russian		Setswana	
	Term	% frequency	Term	% frequency	Term	% frequency
BV-hue	blue	63.8	sinij	44.2	botala	25.0
BV-S2	blue	61.7	seryj-sinij	29.9	botala	5.0
BVB-hue	blue	74.5	sinij	37.7	botala	37.5
B-hue	blue	83.0	sinij	48.1	botala	62.5
BGB-hue	blue	85.1	sinij	49.4	botala	60.0
B-T1	blue	91.5	goluboj	42.9	botala	62.5
BGB-T3	blue	83.0	goluboj	50.6	botala	50.0

BG-T1	blue	63.8	goluboj	33.8	botala	47.5
GBG-S2	blue	59.6	goluboj	27.3	botala	52.5
BG-hue	blue	61.7	morskoj-v.	27.3	botala	50.0
BG-S2	green	59.6	morskoj-v.	26.0	botala	37.5
GBG-hue	green	91.5	zelenyj	49.4	botala	72.5
G-hue	green	91.5	zelenyj	64.9	botala	82.5
GYG-hue	green	89.4	zelenyj	53.2	botala	82.5
YG-hue	green	93.6	zelenyj	61.0	botala	80.0
YGY-hue	green	87.2	zelenyj	51.9	botala	77.5
GYG-T4	green	80.9	zelenyj	28.6	botala	40.0
G-S3	green	89.4	zelenyj	39.0	botala	32.5
YGY-S3	green	85.1	salatovyj	28.6	bosetlha	7.5
YG-S3	green	55.3	zelenyj	40.4	botuba	15.0
Y-S2	green	61.7	xaki	29.9	botuba	10.0
YOY-S2	green	55.3	xaki	15.6	botuba	15.0

Setswana glosses: *botala* 'grue', *botuba* 'pale', *bosetlha* 'yellow'.

Russian glosses: *goluboj* 'light blue', *morskoj-volny* 'sea wave', *salatovyj* 'salad', *sinij* 'dark blue', *seryj-sinij* 'grey-blue', *xaki* 'khaki', *zelenyj* 'green'.

