Color categories: Evidence for the cultural relativity hypothesis[☆]

Debi Roberson[a,*], Jules Davidoff[b], Ian R.L. Davies[c],
Laura R. Shapiro[b]

[a] Department of Psychology, University of Essex, Wivenhoe Park, Colchester, CO3 4SQ, UK
[b] Goldsmiths College, University of London, UK
[c] University of Surrey, UK

Abstract

The question of whether language affects our categorization of perceptual continua is of particular interest for the domain of color where constraints on categorization have been proposed both within the visual system and in the visual environment. Recent research (Roberson, Davies, & Davidoff, 2000; Roberson et al., in press) found substantial evidence of cognitive color differences between different language communities, but concerns remained as to how representative might be a tiny, extremely remote community. The present study replicates and extends previous findings using additional paradigms among a larger community in a different visual environment. Adult semi-nomadic tribesmen in Southern Africa carried out similarity judgments, short-term memory and long-term learning tasks. They showed different cognitive organization of color to both English and another language with the five color terms. Moreover, Categorical Perception effects were found to differ even between languages with broadly similar color categories. The results provide further evidence of the tight relationship between language and cognition.
© 2004 Elsevier Inc. All rights reserved.

---

1. Introduction

The question of whether the language available to describe perceptual experience can influence the experience itself is one that continues to engender lively debate (Boroditsky, 2001; Guest & Van Laar, 2002; Özgen & Davies, 2002; Saunders & van Brakel, 2002). Historically, this debate was characterized by the dichotomous views that thought is either shaped by language (Brown, 1976; Ray, 1952) or completely independent of it (Berlin & Kay, 1969; Heider & Olivier, 1972). Recent systematic investigations of the relationship between language and thought have likewise provided evidence for both views. Differences between languages in grammatical structure and range of terminology have been associated with altered perceived similarity between objects and actions, as well as to different memories of the same experience in the following domains: number systems (Gumperz & Levinson, 1997); spatial relations (Bowerman & Choi, 2001; Levinson, 1996), artifact categories (Malt & Johnson, 1998); modes of motion (Gennari, Sloman, Malt, & Fitch, 2000); time (Boroditsky, 2001); material and shape classification (Lucy, 1992); shape (Roberson, Davidoff, & Shapiro, 2002) and grammatical gender (Boroditsky, in press; Clarke, Lossoff, McCracken, & Still, 1981, 1984; Sera, Berge, & del Castillo Pintado, 1994; Sera et al., 2002). Other studies have argued against the influence of linguistic differences on perceptual classification, both at the level of terminology (Malt, Sloman, Gennari, Shi, & Wang, 1999; Munnich & Landau, 2003) and grammatical structure (Karmiloff-Smith, 1979; Pérez-Pereira, 1991). The present study seeks to shed light on whether language and cognition are coupled or separable in the domain of color categorization and perception.

The field of color categorization has provided a rich testing ground for the effects of language on perception. While the physiological basis of color vision is the same for all humans with normal trichromatic color vision (Jordan & Mollon, 1997), there is considerable diversity in the way that different languages segment the continuum of visible colors. Some languages have been reported to use as few as two terms to describe all visible colors (Rosch Heider, 1972). Others have been reported to use between three and eleven (Berlin & Kay, 1969), while some (e.g., Russian; Davies & Corbett, 1997) may have twelve. This variability exists just for those terms deemed by Berlin and Kay (1969) to be 'basic' (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. However, within these diverse naming systems there are noticeable generalities (Kay, Berlin, & Merrifield, 1991; MacLaury, 1987) It is the finding of such generalities that led to the proposal of panhuman universals in cognitive color categorization that transcend terminological differences (e.g., Heider & Olivier, 1972).

Roberson, Davies, and Davidoff (2000) reported a series of experiments that set out to replicate and extend the work of Rosch Heider in the early 1970s

(Heider & Olivier, 1972; Rosch Heider, 1972). Rosch Heider's experiments had been particularly influential in promoting the view that language and cognitive experience are largely independent (in some cases, orthogonal). Investigating another traditional culture, Roberson et al. found substantial differences in perceptual judgments and memory performance between a language with eleven basic color terms and one with only five (Berinmo). These differences, unlike the data of Rosch Heider from Dani speakers, suggested that language not only facilitates memory performance, but also affects the perceived similarity of perceptual stimuli; a result also found in other cross-cultural investigations of color (Kay & Kempton, 1984; Stefflre, Castillo Vales, & Morley, 1966).

In the present study we provide a further way of examining linguistic relativity. If there is a regular pattern of 'evolution' in color terminologies, as suggested by Kay et al. (1991), then two languages at the same evolutionary 'stage' (having five color terms) might be expected to have similar cognitive representations of color despite having very different environments. However, unlike the population tested by Roberson et al. (2000) this investigation reports findings from Himba participants from twelve villages within a much larger population (estimated between 20,000 and 50,000, Namibian Government statistics, 2004) whose territory is spread over an area of some twenty-five thousand square miles, in northern Namibia, in Africa. Moreover, the Himba are semi-nomadic tribesmen inhabiting an arid region; their visual diet of open desert, scrubland and mountain is radically different to that of Berinmo speakers' deeply shaded and lush forest territory. See Crandall (2000) for an account of the Himba as a distinct, cohesive cultural and linguistic group.

Himba is a dialect of the Herero language, but cultural isolation over the last hundred years has resulted in a variety of cultural and linguistic differences from Herero. The Herero culture is stable and broadly agricultural. Most villages now have schools and radios, and the people have adopted Western dress. Herero has acquired borrowed color terms such as 'grine' and 'pinge' (green and pink) that Himba speakers do not use. Himba people have a strong and distinctive traditional cultural identity. They have retained traditional clothing and lifestyles that bring little contact with other cultures. Himba has five basic terms according to the criteria of Kay et al. (1991), a similar number to the Berinmo language studied by Roberson et al. (2000).

In summary, the aims of this study were to confirm and extend the previous findings, concerning cognitive differences linked to labeling differences, in a population with a similar number of color terms but now with a different visual diet. It also investigated the possibility of differences in the cognitive organization of color between speakers of languages despite the substantially similar sets of color terms.

2. Name Elicitation: Himba basic color terms for saturated stimuli

Himba basic color terms have not been previously established by other researchers (for example, in the World Color Survey of Kay et al., 1991) and so we used the same method of elicitation as Roberson et al. (2000) and Heider and Olivier (1972).

2.1. Method

2.1.1. Participants

Participants in all experiments were monolingual speakers of Himba tested in northern Namibia by the experimenter through the services of a Himba-speaking Herero interpreter, who was naïve to the purpose of the experiments. All participants were screened for color vision defects with both the nonverbal plates of the Ishihara (1992) test for color blindness and the City Colour Vision Test (Fletcher, 1980). All participants were paid in kind. Thirty-one Himba adults (23 females, 8 males) aged between approx. 17 and 55 years carried out color naming for a range of Munsell samples varying in hue, lightness, and saturation.

2.1.2. Materials

The stimuli were identical to those used by Heider and Olivier (1972) and Roberson et al. (2000) and used one hundred and sixty fully saturated, Munsell color chips varying in hue and lightness. The set consisted of hue levels 5 and 10 of ten equally spaced steps around the Munsell circle (Munsell dimension Hue R, YR, Y, YG, G, BG, B, PB, P, RP) each at eight lightness levels (Munsell dimension Value 9/, 8/, 7/, 6/, 5/, 4/, 3/, 2/). Stimuli used for elicitation were 1 in. square glossy finish chips individually mounted on 2 in. square pieces of white card. A second set of chips were mounted on a sheet of white card in their Munsell order (hues horizontally, lightness vertically) for elicitation of best examples of categories.

2.1.3. Apparatus

A solar-powered portable light-box, identical to that described in Roberson et al. (2000) was used for naming tasks. It yielded illuminant C (6700 K) under which Munsell colors are standardized. This approximates to shaded natural daylight, which typically ranges from 5500 to 7500 K as measured by a Gossen Colormaster. A detailed specification is given in Appendix A. A micro cassette recorder was used to record responses.

2.1.4. Procedure

Participants were shown the individual stimuli, one at a time, in random order and asked, in Himba "what color is this?" Himba descriptions of each chip were recorded in full. After completion of the naming task, participants were shown the array containing all the stimulus chips and asked, for each term they had used, to indicate "which of these is the best example of the color......, which is a true one?"

2.2. Results

Fig. 1 shows the modal naming data for basic terms only for Himba participants for comparison with Berinmo and English patterns (reported in Roberson et al., 2000). Himba has five basic color terms (although one, *burou*, is a recently borrowed term from Herero). These are monolexemic, not subsumed under the meaning of other terms, not restricted to a narrow class of objects and understood by all observ-
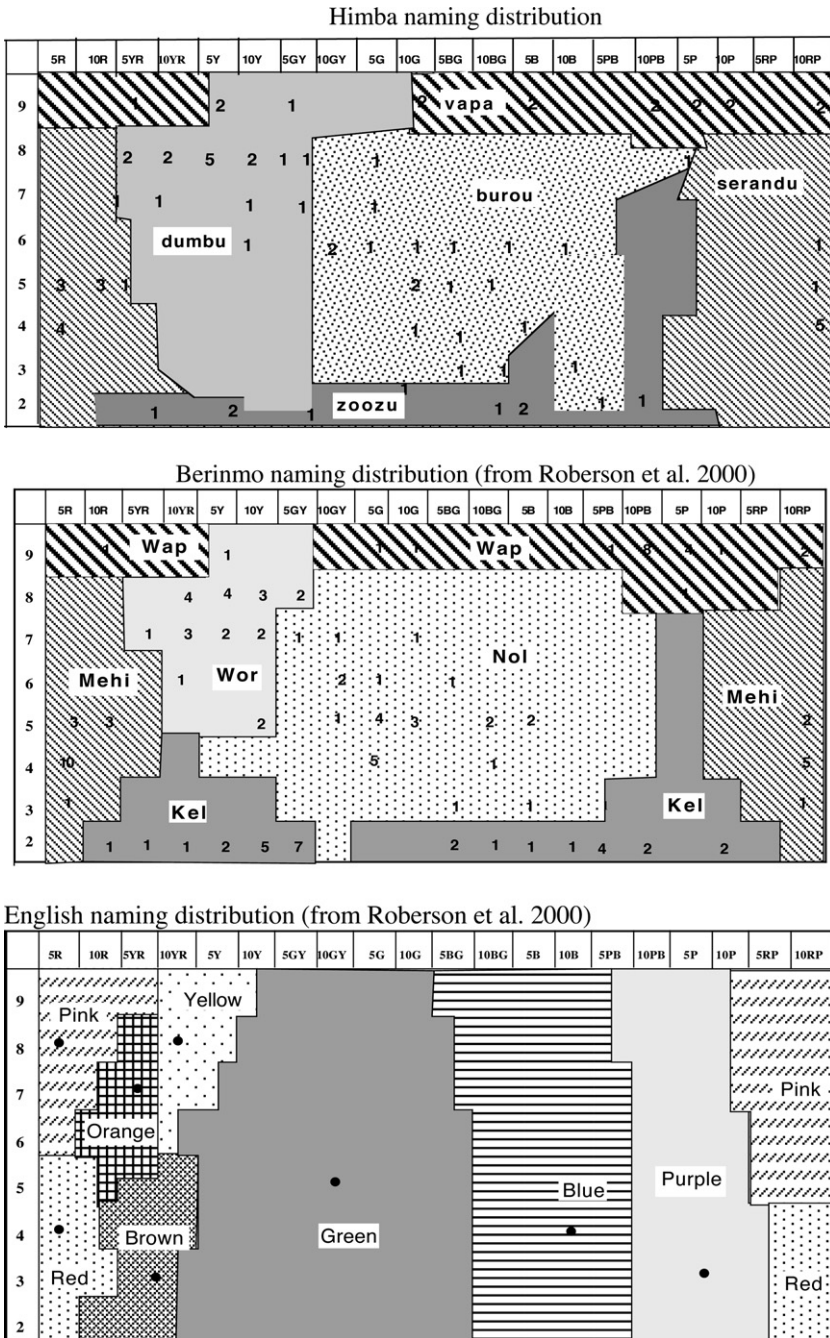
Fig. 1. Distribution of Himba naming and choices of best exemplar for the 160 chip saturated array (for 31 observers) compared to those of English and Berinmo speakers for the same array. Numbers represent number of individuals choosing an exemplar as best example of the category.

ers. Together these terms were used to name 86.2% of all stimuli, (data pooled across participants) compared to 89.2% for Berinmo basic terms. The graph is based on modal naming. Boundaries drawn through an individual chip represent the proportion of each name given to that chip.

A few areas have very low agreement on naming and a few chips have as little as 20% agreement. There are two such areas: one corresponds roughly to English *brown*, the other to English *purple*. In these areas, chips are either named with a basic term; a secondary term; a combination of terms or left un-named. 2.3% of all responses were "don't know"; larger than the 0.89% by Berinmo speakers but similar to the African population tested by Davies and Corbett (1997). 29% of all stimuli were named with greater than 90% agreement. 86% of all chips were named using a single basic term. 8.6% of all names given were secondary terms specific to cattle-hide colors; 2.7% were double terms (e.g., *serandu/vapa*). Modifiers were seldom used. '*Katiti*' (a little) was used for only 0.1% of names. Within-language naming agreement for Himba speakers was .73, compared to .83 for Berinmo speakers. The 5 terms that appear to fit best the criteria for 'basic' are:

Serandu. The Herero/English dictionary (Booysen, 1987) translates this as *red*, but the Himba range of use is quite broad. Used by Himba participants for 20.0% of total naming and to name 41% of total range of chips.

Dumbu. The Herero dictionary translations vary (*beige*, *yellow*). It is also the term for a white person. Used for 17.5% of total naming. Used to name 59% of the total range of chips.

Zoozu. The Herero dictionary translates this as *black*. Only 35% of Himba observers chose a *black* chip as best example; other choices included light *blue* (10 B 6/10), medium *green* (5G 5/10) and best example *blue* (5B 4/10). 2.6% of observers used the term only for the *black* chip. All the rest used it for chromatic stimuli. Used for 9.2% of total naming. Used to name 58% of the total range of chips.

Vapa. The Herero dictionary translates this as *white* but only 35% of Himba observers chose the *white* tile as best example (although all other choices were at the lightest level on the Munsell chart). 22.4% of observers used the term only for the *white* chip. All others extended it to chromatic stimuli. Used for 6.8% of total naming. Used to name 30.8% of total range of chips.

Burou. This term is more recently borrowed from Afrikaans (*blau*) via Herero. The Herero dictionary translates it as *blue*. 16% of observers used it only in with another term for any Munsell stimuli; 6% failed to use it at all.[1] Used for 32.7% of total naming. Used to name 57% of the total range of chips, mostly in the *blue/green/purple* range. This corresponds to a *grue* term in Berlin and Kay's (1969) stage theory. Many surrounding languages have a similar term, although some have separate terms for blue and green.

In addition, a number of secondary terms particular to the Himba dialect and normally used specifically to describe the color of animal hides (cattle, goats, etc.) were

---

[1] Not all observers spontaneously used the term '*Burou*' in the naming task, but those who did not were able to select chips from the appropriate range when asked to indicate a 'Burou' stimulus.

also used by a number of observers (*vinde*, *vahe*, *kuze*, *honi*). These represented 8.6% of total names given. Choices of best examples were quite diverse, but this lack of agreement between different speakers was also found among Berinmo speakers (Roberson et al., 2000), Dani speakers (Heider & Olivier, 1972; in MacLaury, 1987) and for many non-industrialized South American languages (MacLaury, 1987).

2.3. Discussion

Despite the difference in geographical location, the number and range of Himba color terms appeared similar to those of Berinmo. However, the Himba show somewhat less consistency in their naming responses. Three factors may have contributed to the rather less homogenous use of color terms by Himba speakers. First, their solitary cattle herding lifestyle dictates large territories and little cooperative activity; this contrasts with the dense tropical rain forest of the Berinmo people, well stocked with game and vegetation, where communal hunting and gathering activities reinforce tightly knit and inter-related social groups. So, both opportunities and the need to communicate may differ. Second, the Himba society revolves around cattle, whose meat, hide, bones, horns and dung supply food, clothing, tools, decorations and building materials; this appears to have led to the development of a rich vocabulary of secondary color terms that are frequently used, although with little consensus. Last, while Berinmo speakers seemed to relish the challenge of providing a descriptive for every color and object shown to them, Himba people appeared more reticent and thoughtful and were content to leave stimuli un-named. The following experiments attempted to probe the extent of the similarities and differences between the Himba and Berinmo color terms, as well as to compare their cognitive organization to that of English speakers.

3. Experiment 1. Naming and memory for stimuli at low saturation

This experiment was carried out to attempt to further investigate the suggestion of Heider and Olivier (1972) that the underlying cognitive representations of color in speakers of a language with few color terms resembled those of English speakers, rather than being based on their own color names. Heider and Olivier's findings were ambiguous, because their interpretation, based on the visual representation of the scaled data, partly disagreed with the statistical measure of fit. Although the memory patterns of American and Dani speakers were more similar than the naming patterns of the two languages, they were less statistically similar than the Dani patterns of naming and memory. Roberson et al. (2000) found consistently, in opposition to Heider and Olivier, that patterns of memory confusion best matched naming patterns in each language. In Experiment 1, participants were asked to name and remember forty stimuli at low levels of saturation. The conditions of the memory experiment (30 s delay) and the use of very low saturation stimuli were designed to produce high error rates. The present experiment set out to compare those findings to those in the Himba population.

3.1. Method

3.1.1. Participants

Twenty-two adult Himba speakers (8 males, 14 females) aged between approximately 17 and 50 years with normal color vision carried out this task. All were tested by the Experimenter in their home villages through the services of an interpreter.

3.1.2. Materials

The stimuli were identical to those used by Heider and Olivier (1972). Forty Munsell color chips of glossy finish made up of ten hues (hue level 7.5) at four lightness levels (Value 9/, 7/, 5/, 3/) evenly spaced around the Munsell hue circle and all at the lowest possible saturation (/2). The chips measured 1 in. square and were mounted, in their Munsell order (lightness vertically, hue horizontally) on a white A4 board. A second identical set was mounted individually on 2-in. square white card.

3.1.3. Apparatus

The light box was used for both naming and memory tasks.

3.1.4. Procedure

All participants completed both the naming and memory tasks. Half the participants completed the naming task first, while the remainder completed the memory task first. For the naming task, the experimenter displayed the individual chips in the light box, one at a time, and the participant was "What color is this?" All responses were recorded in full. For the memory task, the full array was laid in the light box and covered with a sheet of gray card. Each participant was shown individual test chips, placed on top of the covered array, for 5 s (timed by stopwatch). The chip was then removed and, after a 30-s unfilled delay, the array was uncovered and the participant was asked to select the chip they had just seen. All 40 of the chips were presented to each participant in random order. Participants received no feedback on their performance.

3.2. Results

Basic terms were used to name a mean of 20.77 (52%) of the stimuli in the desaturated array, compared to 38.25 (95%) named with the 5 basic Berinmo terms (Roberson et al., 2000), and 80% named with the 2 basic Dani terms (Heider & Olivier, 1972). Himba participants used significantly fewer basic terms to label the desaturated stimuli than had Berinmo participants. The mean difference between the two was 17.48 ± 4. Of the total number of stimuli *not* named with basic terms, 37.8% were named with secondary cattle or goatskin terms and 10.3% were left un-named. The proportion of un-named stimuli was markedly more than the 1.3% un-named by Berinmo participants, but similar to the Setswana speaking population reported by Davies and Corbett (1997). Table 1 shows the mean correct identifications in the memory test compared to Berinmo speakers (Roberson et al., 2000) and Dani speakers (Heider & Olivier, 1972). In light of the difference in naming behavior of Himba

Table 1
Mean number of correct identifications, (*M*) standard errors (*SE*) and range for Himba speakers in Experiment 1 compared to Berinmo and English speakers (Roberson et al., 2000) and Dani speakers (Heider and Olivier, 1972)

	Himba	Berinmo	English	Dani
M correct identifications out of 40				
M	7.0	9.6	19.9	7.7
SE	1	.8	.7	
Range	(1–13)	(4–17)	(15–25)	(0–17)

participants to previous data, the relationship between naming and memory was further explored by comparing the number of correct identifications to the number of unique descriptors generated. Himba speakers generated a mean of 8.1 unique descriptors over the course of naming all 40 stimuli, compared to 6.9 for Berinmo speakers and 27.6 for English speakers.

For Himba speakers, there was a significant correlation between the number of correct identifications and the number of unique descriptors generated: $r(22) = .48$, $p < .05$. A re-examination of previous data from Roberson et al. (2000) revealed that for Berinmo speakers there was also a significant correlation between the number of correct identifications and the number of unique descriptors generated: $r(22) = .47$, $p < .05$, but for English speakers this relationship was not significant: $r(22) = .08$, $p > .1$. So, at least for languages with few color terms, memory performance is related to the speaker's color vocabulary.

Naming and memory confusions were compared (in dissimilarity matrix form) in replicated multi-dimensional scaling (MDS). Separate 40×40 matrices were constructed both for the memory confusions and the naming data for each participant. The multi-dimensional scaling technique compares these matrices to assess the number of times that two items were either called by the same name or confused in memory. Data points are located in three-dimensional space so that two items would occupy the same point in space were they always called by the same name. Were they never called the same name, they would be placed as far from each other as possible (allowing for their positions relative to all other data points). Fig. 2 shows comparisons of the four resulting scalings of the 40 chips using S stress (squared stress values) to yield a goodness of fit measure for nonmetric scaling solutions between Himba participants and English and Berinmo speakers. Measures of stress indicate the distance that two solutions would need to be moved so that all points in both matrices coincided: therefore, smaller stress values correspond to a better fit. The vertical axis shows the relationship between naming and memory for each language tested. Thus the best fit between naming and memory within a language is for Dani (Heider & Olivier, 1972) followed by Berinmo (Roberson et al., 2000), then English, then Himba and US English.

The fits between memory matrices for each language tested are shown in the top horizontal line of the graph and those between the naming matrices for each language are shown in the lowest horizontal line of the graph. Overall the fit is better
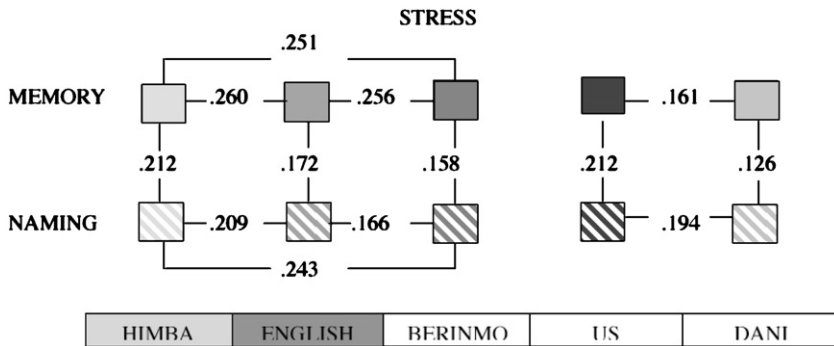
Fig. 2. Stress values for multi-dimensional scaling comparisons of Himba and English naming and memory for desaturated colors, compared to those found for Berinmo and English speakers by Roberson et al. (2000). Note that the higher the score, the poorer the fit.

between naming matrices than between memory matrices, since the fewer terms a language has, the more chips will be labeled with the same term. In all cases except one (US naming and memory) the fit between naming and memory within a language is better than the fit between the memory matrices of corresponding languages. Thus, the statistical fit supports the findings from the Berinmo data. Correlations between the matrices were also calculated, using the Mantel test (Legendre, 2000; Mantel, 1967) for comparison with Roberson et al. (2000) (see Fig. 3A). Differences between the correlations were explored using Fisher's r'. The strength of the correlation between Himba naming and memory ($r = .397$) compared to that between Himba and English memory matrices ($r = .134$) did not quite reach reliability ($z = 1.79$, $p < .08$). However, the relationships between all matrices may be attenuated by perceptual distance, since the most perceptually similar stimuli are more likely both to be called by the same name and to be confused in memory. We therefore calculated the Euclidean distance between each pair of stimuli in the perceptually uniform CIE $L^*u^*v^*$ space (Wyszecki & Stiles, 1982). The matrix data were compared using partial Mantel correlations, controlling for perceptual distance. Fig. 3B shows the partial correlations between matrices. Fisher's r' comparisons of the relationship between matrices when perceptual distance is controlled now revealed that the correlation between Himba naming and memory ($r = .559$) was significantly greater ($z = 3.66$, $p < .001$) than that between Himba and English memory matrices ($r = .036$).

3.3. Discussion

For Himba speakers, as for English and Berinmo speakers, these very desaturated stimuli are poor examples of their basic categories, and thus hard to name. In spite of this, these results (see Figs. 2 and 3) support our previous findings that memory patterns (and presumed cognitive representations) are more similar to patterns of naming within languages than to the memory patterns of other languages. The weakest of the intra-language relationships is that between English naming and memory
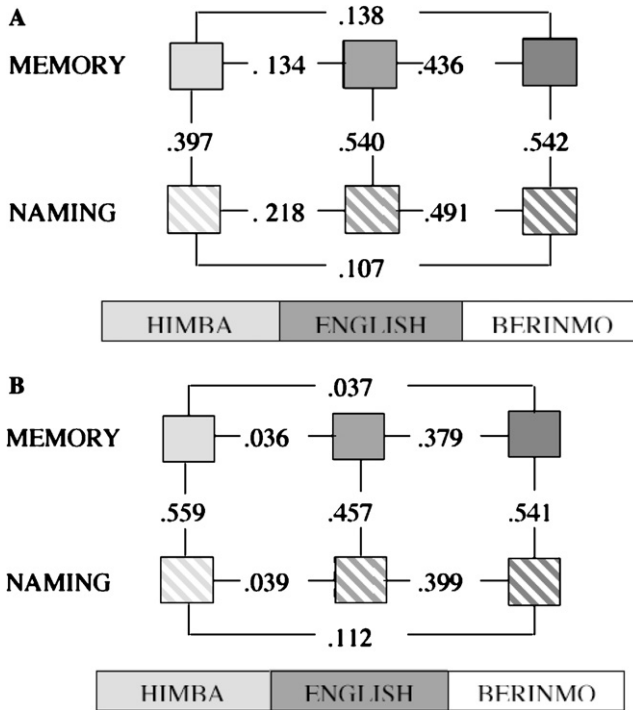
Fig. 3. Correlations between matrices (Mantel test) for naming and memory measures in the three languages tested, before (A) and after (B) the physical distance (measured in $L^*u^*v^*$ space) is partialled out compared to those found for Berinmo and English speakers by Roberson et al. (2000).

($r = .457$ in Fig. 3B) and may reflect the lower error rate (since MDS considers only error data) and the larger number of descriptors used by English speakers. If fewer stimuli are called by the same name, it is less likely that two same name stimuli will be confused in memory. It can also be seen (see Fig. 3A) that there are strong correlations between English and Berinmo naming (.491) and English and Berinmo memory (.436); these reflect the tendency of both groups to make more confusions to other stimuli at the same lightness level. Himba speakers were more inclined to use terms that spanned several levels of lightness. Thus the Himba data give the strongest support to date for the tight relationship between naming and memory.

4. Experiment 2. An examination of the role of focal colors in memory

Rosch Heider suggested, in her seminal studies in the early 1970s, that categories would form around particular 'focal' colors thought to be more perceptually salient, perhaps owing to properties of the visual system (Kay & McDaniel, 1978). Rosch Heider (1972) found that focals for English categories were better recognized and easier to learn than colors that fall outside the center of these basic categories even by

speakers of languages that do not use the English color terms. However, Roberson et al. (2000) failed to replicate these findings, either in short-term memory or long-term learning. Experiments 2a and 2b considered the role of English focal colors in short-term memory and long-term learning for Himba participants.

5. Experiment 2a. Focal vs. non-focal short-term memory

Accuracy scores for English focal, internominal and boundary chips were collected in the same paradigm as had been used for Berinmo and English speakers (Roberson et al., 2000), and previously for Dani speakers (Rosch Heider, 1972). The three sets of stimuli were chosen by Rosch Heider (1972) after extensive pilot research to establish the foci of the basic categories of English.

5.1. Method

5.1.1. Participants
Twenty-four monolingual Himba adults (9 males, 15 females) with normal color vision, between the ages of approximately 17 and 55 were tested by the Interpreter in their home villages, in the presence of the Experimenter.

5.1.2. Materials
The eight best examples of the English basic chromatic categories (*red*, *yellow*, *green*, *blue*, *orange*, *purple*, *pink*, *brown*) together with sixteen color chips that fall outside the focal areas of English categories and are therefore difficult for English speakers to name, were used as targets, following Rosch Heider (1972). Targets designated as 'Boundary' chips by Rosch Heider are from those areas towards the outside of basic categories, those designated as 'Internominal' fall between two basic categories. The full array of 160 stimuli used for naming was used to test recognition. Munsell designations of the target chips were for focal chips: *red* 5R 4/14, *yellow* 10YR 8/14, *green* 10GY 5/12, *blue* 10B 4/10, *pink* 5R 8/6, *orange* 5YR 7/14, *brown* 5YR 3/6, and *purple* 5P 3/10. The internominal chips were: 5YR 8/8, 10Y 6/10, 10Y 4/6, 10GY 8/8, 10BG 7/6, 10BG 4/6, 5P 7/8 and 5RP 3/8. The Boundary chips were: 5R 2/8, 5Y 3/4, 10Y 9/6, 10GY 7/10, 10G 3/4, 5B 6/10, 10B 3/6, and 5P 6/8.

5.1.3. Illumination
All stimuli were tested in conditions of natural daylight for comparison with previous experiments.

5.1.4. Procedure
The order of the two test sheets in the 160-chip array was counterbalanced, following the procedure introduced by MacLaury (1987), as in Roberson et al. (2000). The array was placed in front of each participant and covered. Participants were then shown a single target chip for 5s (timed by stopwatch). The target chip was withdrawn and, after a 30s unfilled interval, the test array was uncovered and participants

were instructed to select the test chip matching the target they had just seen. All participants first received two practice trials with chips that were not in any target category and all received the target chips in random order.

5.2. Results

As with previous data for Berinmo speakers, Himba recognition accuracy was analyzed using the bias-free d' measure to account for guessing. Data were analyzed in a 3 level (target type: focal vs. internominal vs. boundary) repeated measures ANOVA. There was no significant effect of target type [$F(2,46) < 1$]. The Berinmo speakers, reported in Roberson et al. (2000), also showed no significant effect in a similar analysis, [$F < 1$]. Mean d' scores are shown in Table 2.

It might, however, have been the case that the four focal colors deemed 'primary' (Rosch Heider, 1972) *red*, *yellow*, *green* and *blue*, were better recognized than the 'secondary' *colors pink*, *purple*, *orange*, and *brown*. To check this the mean d' score was calculated for primary and secondary colors. For Himba speakers, the mean d' score for primary colors was 1.42, that for secondary colors was 1.58. For Berinmo speakers the mean d' score for primary colors had been 1.86, that for secondary colors was 1.61. There was no reliable advantage for primary over secondary focal colors in either language.

Himba categories have a broad focus and several of the 'non-focal' examples of English categories have focal status in Himba categories, as indexed by naming agreement. Naming agreement was used as an index of focality, since the choice of the original set of English focal tiles by Rosch Heider (1972) was also made on the basis of naming. In particular, there are four tiles that are considered 'internominal' for English speakers (falling between two basic name categories) that receive higher name agreement (over 95%) for a Himba term than their 'focal' equivalents. A further analysis compared performance across Himba participants for stimuli that were either focal to both languages, focal to Himba only or focal to English only. A fully within-subjects ANOVA comparison of d' scores for the three types of stimuli for Himba speakers revealed a significant effect of stimulus type [$F(2,46) = 3.29$, MSe $= .33$, $p < .05$], (see Fig. 4). An examination of the contrasts between means revealed that, while there was no difference between the number of stimuli recognized that were focal only in Himba or in both languages, those stimuli that were focal in Himba only were recognized significantly more often than those focal only in English ($p < .05$).

A re-examination of the Berinmo data (Roberson et al., 2000) also found some overlap in focality, based on 95% naming agreement for the stimuli. Again, there

Table 2
Mean d' recognition scores for Himba speakers, compared to English and Berinmo speakers (from Roberson et al., 2000), for focal and non-focal chips

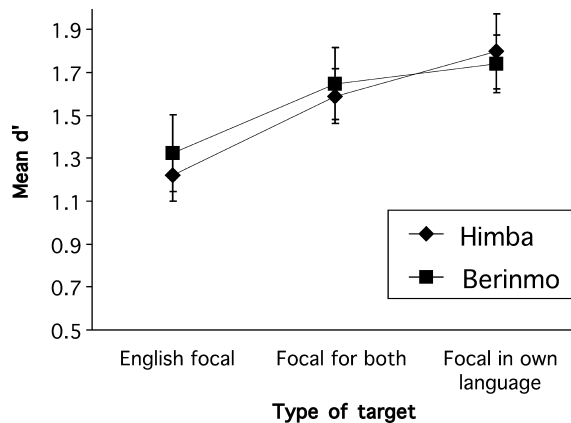| Stimulus type	Himba speakers	English speakers	Berinmo speakers
Focal	1.50 (.23)	2.47 (.24)	1.61 (.14)
Internominal	1.58 (.15)	1.92 (.13)	1.73 (.11)
Boundary	1.80 (.07)	1.56 (.22)	1.62 (.20)

Standard errors in brackets.

Fig. 4. Mean d' correct identifications and standard errors for targets focal in English, in Himba or in both languages in Experiment 2a, compared to those for Berinmo speakers (from Roberson et al., 2000) for targets focal in English, Berinmo or in both languages.

were some stimuli that are non-focal for English speakers that receive higher name agreement for a Berinmo term than their 'focal' equivalents. A similar fully within subjects ANOVA comparison of d' scores for the three types of stimuli for Berinmo speakers now revealed a significant effect of stimulus [$F(2,60) = 3.18$, MSe $= .58$, $p < .05$]. An examination of the contrasts between means revealed that, while there was no difference between the number of stimuli recognized that were focal only in Berinmo or in both languages, those stimuli that were focal in Berinmo only were recognized significantly more often than those focal only in English ($p < .05$). Fig. 4 illustrates these results.

5.3. Discussion

Experiment 2a found no evidence that English focal colors are important for the Himba. Himba performance was poor overall, but at a comparable level to that found for Dani speakers (Heider & Olivier, 1972). Furthermore, when comparisons were carried out, for both the present data set and that previously collected from Berinmo participants, both languages showed superior recognition memory for those stimuli that were most consistently named (focal) in their own language. The superior recognition cannot be an artifact of greater perceptual distinctiveness or to the position of targets in the test array, since the critical stimuli differed across languages. What appears to be universal, in this case, is the tight link between naming and memory.

6. Experiment 2b. Long-term learning of focal vs. non-focal targets

The methodology followed that used for Berinmo speakers (Roberson et al., 2000). Pictures were used as paired associates to colors with each participant learning only 8 (4 focal, 4 non-focal) pairs. However, in the case of Himba participants they

learned paired associates to pictures of Namibian cattle rather than to pictures of palm nuts.

6.1. Method

6.1.1. Participants

Twelve monolingual Himba adults (4 men, 8 women) with normal color vision, between the ages of approximately 17 and 40 were tested by the Experimenter in their home villages, through an interpreter.

6.1.2. Materials

The eight focal and eight internominal chips used in Experiment 2a were used in Experiment 2b. Chips measured 20×20 mm and were mounted in the center of 50 mm square pieces of off-white card. Sixteen pictures of Namibian cows were created from photographs of local cattle, adjusted so that the outline shape was an identical side view, but coloring, shading and pattern varied individually. All cattle stimuli were shown against a uniform blue background.

6.1.3. Illumination

All stimuli were displayed in the light box used for naming.

6.1.4. Procedure

Participants were required to learn eight stimulus-response pairs. Each of the eight target stimuli was paired with a separate response picture. The task was described as learning a new game that the Experimenter would teach them and for which they would be rewarded when training was complete. The eight cattle pictures were first laid out in random order in the light box. The Experimenter showed the participants each of the color stimuli as a paired associate to each picture. The stimuli were then shuffled and the participant was presented with the stimuli, one at a time and asked to pair them with the appropriate picture. Feedback was given after each response and participants were corrected if they made an error. Then the target stimulus was removed from the box before presentation of the next target. Participants completed five training sessions per day for five days, or until the criterion of one perfect run had been achieved. A full record was kept of participants' responses on each run.

6.2. Results

All 12 participants reached criterion within 5 days of training. The mean number of trials to criterion for the full set of items was 20.5 (4.1 days) as opposed to 16.4 (3.3 days) for Berinmo speakers. Mean errors out of 25 for focal and non-focal items are shown in Table 3 for Himba and Berinmo speakers.

There was no advantage for focal over non-focal items. The mean difference between number of errors was $.17 \pm 1.98$ (a similar lack of difference was found for Berinmo speakers in Roberson et al., 2000).

Table 3
Mean errors to criterion for Himba and Berinmo speakers (from Roberson et al., 2000) learning paired-associate pictures to focal and non-focal colors (standard errors in brackets)

	Focal colors	Non-focal colors
Himba speakers	14.46 (.57)	14.59 (.73)
Berinmo speakers	8.31 (.86)	8.15 (.11)

To assess whether *some* participants nevertheless did learn paired-associates to the focal targets with fewer errors than to non-focals, the data were also considered by participant. Again there was no advantage for focal items. The mean difference between number of errors was $.59 \pm .39$; this is also in line with the finding for Berinmo speakers.

6.2.1. Order of learning

The order in which focal colors were learnt by Himba speakers, from lowest to highest errors, (with mean number of errors/25 in parentheses), was: *yellow (9.83)*, *red (13)*, *orange (14)*, *purple (14.17)*, *green (15.17)*, *pink (15.5)*, *blue (15.67)*, *brown (16.17)*. For Berinmo speakers it had been: *red/pink (4.83)*, *purple (7.83)*, *green (8.67)*, *blue (9)*, *brown/orange (9.83)*, *yellow (11.67)*.

6.3. Discussion

No evidence for an advantage for 'focal' items in long-term learning was found in this experiment. It was not possible to carry out the comparison for stimuli that were focal in each language, as done in Experiment 2a, since participants were asked to learn only four of the English focals and four non-focals. Thus, there was insufficient data for formal analysis. Overall Himba participants made more errors across learning trials than Berinmo participants; this may have been because the pictures of cattle used as paired associates were less distinctive than the set of palm nuts used with Berinmo speakers. However, the results of Experiment 2 clearly support the view that language and cognitive representation for color are tightly linked. We continued our investigation by looking for evidence of Categorical Perception at several different category boundaries.

7. Experiment 3. An examination of categorical perception

Himba speakers were tested using three paradigms previously used to assess Categorical Perception (CP) in Berinmo speakers. The aim was to see if the Himba showed any tendency towards CP for *blue* and *green*, or whether, like Berinmo speakers, they would show it only at the boundary of their own name categories. Harnad (1987) provides a comprehensive discussion of CP across a range of auditory and visual categories. CP, in the color domain, means that the physical continuum of the chromatic spectrum is perceived as qualitatively discontinuous, discrete

segments (*red*, *orange*, *yellow*, etc.) Items from different categories (relative to an objective measure such as a just-noticeable-difference) are judged as greater than distances between items from the same category (Bornstein, 1987). CP for color is manifest by faster and more accurate discrimination of two colors that cross a category boundary (e.g., between *green* and *blue*) than of two colors that are both good exemplars of the same category. When testing Berinmo participants, Roberson et al. (2000) compared the boundary between *blue* and *green* (in English) with the boundary between the Berinmo terms *nol* (corresponding to the English categories: *green + blue + purple*) and *wor* (corresponding to the English categories: *yellow + green + orange + brown*). For both pairs of categories, it was possible to construct a set of stimuli crossing the boundary that varied only in hue, maintaining lightness and saturation constant. Himba participants were tested in each paradigm on one English boundary (between *blue* and *green*), one Berinmo boundary (between *nol* and *wor*) and one Himba boundary (between *dumbu* and *burou*) for comparison with previous data.

The Himba boundary (between *dumbu* and *burou*) provides an opportunity both to directly compare Himba and English linguistic boundaries (as was done for Berinmo). It was also possible to compare Himba speakers CP with that of Berinmo speakers since both systems have five basic color terms and there is some similarity between the range referents for each term in each language. In the case of the two Himba categories, *dumbu* and *burou* and the two Berinmo categories *nol* and *wor*, the position of the boundary differs only by a small amount. Thus, it is possible to compare performance across the two boundaries within the same set of stimuli. The following three experiments examine CP for colors in judgments of similarity, in two-forced-choice alternative memory judgments and in a category-learning paradigm.

8. Experiment 3a: Similarity judgments

Sets of triads of color stimuli were created ranging either across the boundary between the English categories of *blue–green*, the Berinmo categories of *nol–wor*, the Himba categories of *dumbu–burou*. Following Roberson et al. (2000), Himba speakers were asked to make similarity judgments for triads of stimuli ranging across each boundary.

The prediction, based on findings with Berinmo speakers, was that Himba speakers would judge within-category stimuli to be more similar to each other than cross-category stimuli where their own linguistic boundary coincided with the boundary of the set, but not otherwise. The Himba term *burou* is a recently borrowed one from Herero, which is still not used with consistency by some Himba people. As the sets of stimuli used included some intermediate Munsell steps, not named in the initial naming phase of these experiments, a pilot study was conducted in which 48 young adults (17 males, 31 females) with normal color vision were asked to name the stimuli in each range, in random order. Appendix B shows the percentage of name agreement within each language for the relevant stimuli.

On completion of the similarity judgment tasks in the present experiment, Himba participants were also asked to name each of the stimuli in each set. Data was excluded for two older participants who failed to use the term *burou* to name any of the stimuli in the sets.

8.1. Method

8.1.1. Participants

Twelve Himba adults (4 males, 8 females), screened for normal color vision, judged the *blue–green* set and the *nol–wor / dumbu–burou* set. All were paid, in kind, for their participation in these experiments.

8.1.2. Materials and Apparatus

The stimuli were sets of individual glossy Munsell chips at the same level of lightness and saturation, surrounding the boundary chip for each set. The 9 chips in the *blue–green* and *nol–wor / dumbu–burou* sets all had Value (lightness) level 5 and Chroma (saturation) level 8.

Each triad of stimuli was displayed, mounted in an equilateral triangle on a piece of off-white Munsell display card measuring $80\,mm \times 100\,mm$ and the relative position of the three chips in each triad was counterbalanced across presentations. For each set, the nine triads were composed as follows: two triads were fully within one category; two triads crossed the category boundary with two chips in one category and one in the other; in two triads the boundary chip was central and three triads the boundary chip was peripheral. Table 4 shows the composition of equivalent triads in each of the three sets. It was predicted for each language that, where all three chips were fully within a category, the two furthest from the boundary would reliably be judged most similar. This was also predicted for triads where the boundary chip was peripheral. Where two stimuli were within one category and the other lay across the boundary, the two within-category chips should be chosen as most similar. Finally, for the two triads in which the boundary chip was central, no specific pairing was predicted. These triads were included because their status was different for the other language under consideration.

8.1.3. Procedure

Each participant was seated at a table, in front of the light box and the triads in a set were presented, one at a time, in random order. Participants were asked to judge "which two of these three colors look most like each other, in the way that brothers look like each other?" (following Roberson et al., 2000). To ensure that participants understood the task, they first completed two practice trials with stimuli that were not part of any set, and for which two stimuli were clearly close and the third very distant (e.g., two *purple* and one *brown*). For those participants judging the *blue–green* and *nol–wor / dumbu–burou* sets, order of sets was counterbalanced and there was a one week interval between testing for the two sets. Participants indicated their responses by pointing. Each of the nine triads in each set was repeated four times varying the position of stimuli on the display card.

Table 4
Composition of triads used for judgments of similarity in Experiment 3a

	7.5 B	5B	2.5B	10BG	7.5BG	5BG	2.5BG	10G	7.5G
blue – green	7.5 B	5B	2.5B	10BG	7.5BG	5BG	2.5BG	10G	7.5G
nol - wor	5Y	7.5Y	10Y	2.5GY	5GY	7.5GY	10GY	2.5G	5G
dumbu - burou	5Y	7.5Y	10Y	2.5GY	5GY	7.5GY	10GY	2.5G	5G
Triad 1		*	*	*					
Triad 2	*		*		*				
Triad 3		*		*		*			
Triad 4				*	*	*			
Triad 5			*		*		*		
Triad 6				*		*		*	
Triad 7					*	*	*		
Triad 8					*		*		*
Triad 9						*	*	*	

Category boundary

* Indicates the chips making up each triad.

8.2. Results

Table 5 shows the mean proportion of predicted choices, by Himba participants, for the seven triads for which specific predictions were made (compared to those obtained from data in Roberson et al., 2000 for Berinmo and English speakers) at each boundary. For the *blue–green* boundary, a between participants ANOVA showed a significant effect of language, $[F(2,25) = 7.81, \mathrm{MSe} = 9.67, p < .01$. A contrast of means revealed that English speakers made significantly more predicted pairings than Himba speakers ($p < .01$) or Berinmo speakers ($p < .05$). For the *nol–wor / dumbu–burou* boundaries, data were analyzed separately for each language, because the predictions for Himba and Berinmo are identical for five of the nine triads (triads 1, 2, 3, 8 and 9), but differ across languages for the other four (triads 4, 5, 6 and 7) since the boundary differs in each languages. English speakers were at chance (less than .5 judgments) for either set of predictions (.45 for *nol–wor*; .47 for *dumbu–burou*).

Table 5
Mean proportion of similarity judgments following predictions based on the category boundary for sets of stimuli crossing *blue–green*, *nol–wor*, and *dumbu–burou*

	Himba	Berinmo	English
blue–green boundary	.54	.65	.72
nol–wor boundary	.56	.79	.45
dumbu–burou boundary	.76	.52	.46

Berinmo and English data from Roberson et al. (2000).

For Himba speakers the mean proportion of predicted choices for the *dumbu–burou* boundary was .76, significantly greater than chance on a binomial test ($p < .05$).

To investigate whether Berinmo and Himba speakers would show differential CP at their respective boundaries, the proportion of predicted pairings chosen for the five identical predictions was compared to those for the two triads where a definite prediction was made only for Himba and the two where a definite prediction was made only for Berinmo in a 2: (Language: Himba vs. Berinmo) × 3 (Prediction: identical for both languages vs. predicted for Himba vs. predicted for Berinmo) ANOVA, with repeated measures over the second factor. There was no significant effect of Language: [$F(1,18) < 1$] but a significant effect of Prediction: [$F(2,36) = 12.52$, MSe = .05, $p < .001$], as well as a significant interaction: [$F(2,36) = 11.07$, $p < .001$]. A Newman–Keuls pairwise comparison of the interaction revealed that Himba speakers made more of the predicted choices for the Himba prediction than Berinmo speakers ($p < .01$) and Berinmo speakers make more of the predicted choices for the Berinmo prediction than Himba speakers ($p < .01$). Fig. 5 illustrates these results.
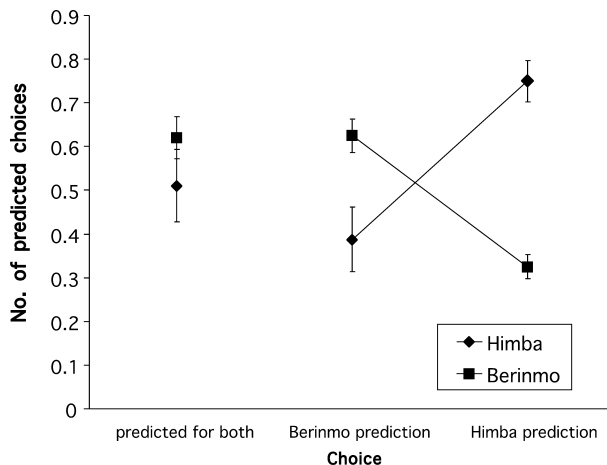

Fig. 5. Mean proportion of predicted judgments and standard errors for triads where the identical judgment is predicted for both languages and triads where the judgment is predicted for Himba, or Berinmo (from Roberson et al., 2000) in Experiment 3a.

One possible reason for the difference in choices of most similar items is that Himba and/or Berinmo speakers misinterpret the task and choose two chips as being more similar only if they have a common name and the third chip has a different name, choosing randomly otherwise. To eliminate this possibility, we examined responses for both groups just for those triads where all three chips would be given the same name with the prediction that the two chips furthest from the relevant boundary would always be chosen as most similar, since they are better examples of the category than the item closest to the boundary. For the *nol–wor / dumbu–burou* boundary, Himba speakers made .79 predicted choices for triads where all three chips would have the same name in Himba. Berinmo speakers made .77 predicted choices for triads where all three chips would have the same name in Berinmo. Both results are significantly different from chance on a binomial test ($ps < .01$).

8.3. Discussion

An argument against the recent behavioral evidence in favor of the linguistic relativity position is that so much of the data depends on judgments in memory where labels could become critical. Furthermore, recent investigations of categorical perception (Özgen & Davies, 2002; Roberson & Davidoff, 2000) have produced results favoring a verbal locus for perceptual memory judgments. Roberson and Davidoff (2000) found that effects attributed to categorical perception (better cross-category than within-category discrimination) disappeared under verbal interference and Özgen and Davies (2002) found that novel category boundaries may be established by common labels (see also Goldstone, 1994, 1998; Özgen, 2004). These results might suggest that categorical perception is based on a cross-category advantage in labeling and not a genuinely perceptual phenomenon. Thus, evidence in Experiment 3a with similarity judgments that do not require stimuli to be remembered is particularly important.

Himba participants, like English or Berinmo speakers, consistently judge stimuli to be more similar if they come from within the same category than if they come from different categories. Moreover, despite the similarity between Himba and Berinmo categories, when we consider only those triads for which the boundary differs between the two languages, the effects of linguistic category are quite specific. These category effects are not artifacts from misunderstanding the task since, even for those triads where all three chips have the same name, the two chips furthest from the boundary are consistently chosen as most similar. Rather, these data are more sympathetic to a perceptual basis for effects of categorical perception (Pilling, Wiggett, Özgen, & Davies, 2003; Roberson, Davidoff, & Braisby, 1999). Pilling et al. (2003) showed that, under some circumstances, categorical effects survive verbal interference and Roberson et al. (1999) that, while language impairment (an inability to name colors) prevents explicit color sorting and the use or comprehension of color terms, it does not prevent implicit color categorization or categorical perception of colors.

These studies suggest that a genuinely perceptual difference is being tapped in similarity judgments, as do the cross-lingual data in Experiment 3a. Of course, label learning could drive perceptual learning of categories, as would be possible in the following studies of recognition memory.

9. Experiment 3b: 2-Alternative forced-choice recognition memory judgments

The Himba were also asked to make two-alternative forced-choice memory judgments for pairs of stimuli that either crossed the category boundary or were entirely within one or other category. The paradigm (Roberson et al., 2000) has previously shown that English speakers demonstrate CP by more accurate cross-category than within-category discrimination (Pilling et al., 2003) and that Berinmo speakers show CP for the boundary between *nol* and *wor* (Davidoff, Davies, & Roberson, 1999). As with the similarity judgments, comparisons are made between Himba and Berinmo judgments for the Himba and the Berinmo boundaries.

9.1. Method

9.1.1. Participants
Twelve Himba adults (6 males, 6 females) screened for normal color vision, were paid, in kind, for their participation in this experiment.

9.1.2. Stimuli and apparatus
Four pairs of within-category stimuli and four pairs of cross-category stimuli were created from the stimuli used in Experiment 3a for *blue* and *green* and an additional four pairs of each type for the *nol–wor / dumbu–burou* range. The pairs were constructed so that either both members were within the same category (e.g., 5B–10BG) or the pair lay across a category boundary (e.g., 10BG–5BG). One-step cross-category trials included the boundary chip and the chip on one side or other of it. These pairs can be considered to straddle the boundary inasmuch as the boundary chip is named approximately equally often as either in one category or the other. These pairs have been shown to produce better discrimination than those that are fully within-category (Bornstein & Korda, 1984). A full list of pairs can be found in Appendix C. Lighting and testing conditions were identical to those used in Experiment 3a.

9.1.3. Procedure
Each member of each pair was shown twice as the target stimulus.

Target stimuli were displayed in the light box in front of participants for 5 s and then removed. After a 5 s unfilled interval, the test pair of stimuli were placed in the light box and the participant was instructed to point, as quickly as possible to the chip matching the target. Position of the target in the test pair was counterbalanced and order of presentation of the pairs was randomized.

9.2. Results

Himba adults' recognition of cross-category and within-category pairs of stimuli was examined in a 2 (Boundary: *blue–green* vs. *dumbu–burou*) × 2 (Pair type: Within vs. Cross) fully within-subjects ANOVA. There was a significant effect of Pair Type

Table 6
Mean correct 2-AFC memory judgments for each population based on a category boundary for stimuli crossing the *green–blue* boundary and the *dumbu–burou* boundary

	Himba	Berinmo	English
blue–green cross-category	.59	.66	.74
blue–green within-category	.61	.71	.63
dumbu–burou cross-category	.81	.72	.73
dumbu–burou within-category	.65	.82	.70

Berinmo and English data from Roberson et al. (2000).

$[F(1,11) = 5.96,$ MSe $= 3.5, p < .05]$, but the effect of Boundary failed to reach significance $[F(1,11) = 4.06,$ MSe $= 4.6, p < .07]$. There was also a significant interaction $[F(1,11) = 4.99, p < .05]$. A Newman–Keuls pairwise comparison of the interaction showed that for the *dumbu–burou* boundary, but not for the *blue–green* boundary, Himba speakers showed significantly better recognition of targets from their cross-category than from within-category pairs ($p < .01$). Table 6 shows the proportion of correct choices for cross- and within-category pairs for Himba speakers compared to that for Berinmo and for English speakers.

For the *nol–wor/dumbu–burou* sets of stimuli, a similar analysis was carried out to that in Experiment 3a, comparing those pairs for which predictions for the two languages differed. Previous results (Roberson et al., 2000) had already shown that Berinmo participants recognized significantly more cross- than within-category targets for the *nol–wor* boundary. Of the eight pairs of stimuli, four pairs were either within- or cross-category for both languages. However, for the remaining four pairs, two that were within-category for Berinmo were cross-category for Himba, and vice versa. An analysis was carried out on the recognition accuracy for these critical items.

A two (Language: Himba vs. Berinmo) × two (Target type: Within-category for Himba vs. Cross-category for Himba) mixed design ANOVA with repeated measures over the last factor, revealed no significant effect of language $[F(1,18) < 1]$ and no significant effect of target type $[F(1,18) < 1]$, but a significant interaction $[F(1,18) = 4.56,$ MSe $= .57, p < .05]$. Newman–Keuls pairwise comparison of the interaction (see Fig. 6) revealed that only Himba participants recognized significantly more of the Himba cross-category pairs than the Himba within-category pairs ($p < .05$).

9.3. Discussion

With regard to the English categories of *blue* and *green*, Himba speakers, like Berinmo speakers, fail to show the better discrimination of cross-category pairs that is the hallmark of CP. They do, however show better discrimination of cross-category pairs for the *dumbu–burou* boundary. Moreover, we again observed the difference associated with a slightly shifted boundary for *dumbu–burou* relative to *nol–wor*. For speakers of both languages, the enhanced discrimination of stimuli crossing the category boundary is language specific.
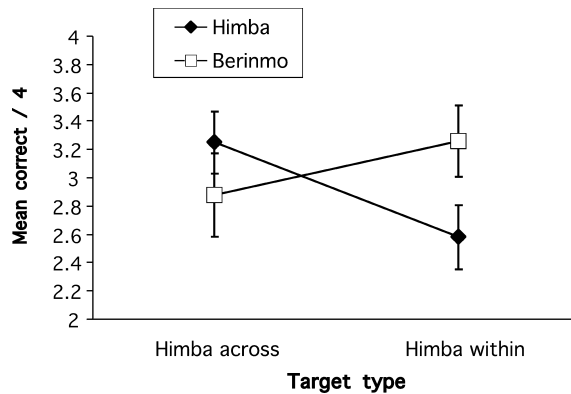
Fig. 6. Mean correct identifications (max = 4) and standard errors for Himba and Berinmo speakers (data from Roberson et al., 2000) for pairs of stimuli that are either within-category or cross-category for the Himba language.

10. Experiment 3c: Category learning for English and Himba color categories

To further explore the hypothesis that categories are learned through language, Himba participants were asked to learn to divide sets of stimuli into two groups according to either English or Himba category boundaries. Following Roberson et al. (2000), learned divisions of sets of stimuli crossing the *blue–green* boundary were compared to the arbitrary division of a set falling fully within the English category *green* (*green* 1 vs. *green* 2). Himba and English participants were also asked to learn to divide a set crossing the English boundary between *yellow* and *green*; this was compared to another set of stimuli, crossing the Himba *dumbu–burou* boundary, equated to the *yellow–green* set for variability of lightness and saturation.

10.1. Method

10.1.1. Participants
Twelve Himba adults learnt the *blue–green* and *green* 1–*green* 2 category divisions. A further twelve learnt the *yellow–green* and *dumbu–burou* divisions. All participants were screened for normal color vision, and all were paid, in kind, for their participation.

10.1.2. Stimuli and apparatus
Sets of stimuli were constructed for each of the four boundaries in the same manner. For example, for the *blue–green* boundary, 10 *blue* stimuli were matched to 10 *green* stimuli in their distance from the boundary (7.5BG) as well as in their range of lightness and saturation. Appendix D contains a full list of the stimuli used in each set. To make the Himba sets non-trivial for Himba speakers, all sets included poor as well as good examples of the categories.

10.1.3. Procedure

Participants were taught, with feedback, to sort sets of stimuli into two categories, following the procedure used in Roberson et al. (2000). Each participant learnt to divide first one, and then another set of stimuli into two categories, with the training sessions for each set separated by at least a week. Order-of-learning the sets was counterbalanced across participants and spatial locations (pointing to the left and right side of the light box) were used rather than verbal responses, to minimize the requirement for linguistic labeling. The experimenter showed each participant three samples of each category (chosen at random from the set), which were placed, one at a time, either to the left or right inside the light box, as the interpreter said "this one goes on this side" for each sample. Sample stimuli were removed and the whole set placed, one at a time, in random order, in the center of the box. Participants were asked to indicate, by pointing, whether each stimulus should go to the right or left. Participants were praised for a correct choice and corrected for an incorrect one, before that stimulus was removed. All responses were recorded and the training session continued to a criterion of one complete correct sort.

10.2. Results

For the *blue–green* and *green* 1–*green* 2 sets; the mean difference between the rates of learning for the two sets was $.83 \pm 4.26$. Table 7 shows mean errors to criterion (and standard errors) for Himba participants, compared to those for Berinmo and English participants reported in Roberson et al. (2000). For the *dumbu–burou* and *yellow–green* divisions, the mean number of trials to criterion for *dumbu–burou* was 4.83 (.59) and for *yellow–green* was 9.5 (1.17). The mean difference between the rates of learning for the two sets was 4.67 ± 2.71, with an advantage for Himba categories.

10.3. Discussion

In a category-learning paradigm, there was no evidence that Himba participants perceived the *blue–green* region of color space in a categorical manner. Like Berinmo

Table 7
Mean errors to criterion for the three populations learning the *blue* vs. *green* and *green* 1 vs. *green* 2 category divisions

	blue–green	green 1–green 2
Himba speakers	10.42 (1.37)	9.58 (1.53)
Berinmo speakers	11.43 (0.97)	10.57 (0.53)
English speakers	3.14 (0.51)	6.29 (0.94)
	yellow–green	dumbu–burou
Himba speakers	9.50 (1.17)	4.38 (0.59)
English speakers	3.61 (0.17)	4.25 (0.22)

Berinmo and English data from Roberson et al. (2000). Below these are data for Himba speakers learning the *yellow–green* and *dumbu–burou* category divisions.

speakers, they did not find this division easier to learn than an arbitrary one in the center of the *green* category. There was also a significant advantage for learning the *dumbu–burou* division, over the *yellow–green* division. It thus appears that CP for color category boundaries is tightly linked to the linguistic categories of the participant.

11. General discussion

There is a growing body of evidence that speakers of different languages, whose terminology or grammatical structure differ, encode, remember and discriminate stimuli in different ways. The present studies addressed the question by asking whether, and to what degree, language influences the particular task of color categorization. The data for the Himba (Experiments 1 and 2) gave the same positive answer to the question, as did our earlier study with the Berinmo of Papua New Guinea (Roberson et al., 2000). Moreover, despite the apparent similarity between the basic linguistic color categories in Himba to those in Berinmo, the differences allowed language-specific category effects. Himba participants show 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 (Experiments 3a and b).

Two essential questions arise from the continuing debate over the existence of a particular set of Universal color categories (Kay & Regier, 2003; Munnich & Landau, 2003; Saunders & van Brakel, 2002). One is quantitative: how different must two sets of descriptive terms be before there are observable cognitive consequences of those differences for speakers of two different languages? The other is qualitative: should categories be defined by their best examples/centroids/foci or by their full range? If the former, then two categories both count as *green* just if they both include the English best example *green* even though one may include colors that would be called *brown*, *blue* and *purple* in English. If the latter, then two categories would count as different just if they include different sets of exemplars. The qualitative issue was raised by Rosch Heider (1972, 1973) with regard to whether categories form around prototypes.

The present findings, in answer to the quantitative question, suggest that quite small differences in boundary position are sufficient to yield observable cognitive differences and that the boundaries between categories are at least as salient as their centers. In consequence, when considering the qualitative question of whether two categories, in different languages, are effectively identical, the full set of exemplars should be considered, rather than just the category centre. Thus, both questions can only be answered in the context of linguistic constraints. An alternative approach, however, would concentrate on answering these questions from the overall similarity between languages' color terms rather than the differences. In particular, overall similarity could argue against linguistic relativity especially as the similarity extends to many of the world's languages (Kay & Regier, 2003; Lindsey & Brown, 2002). We wish to comment on these similarities but it should be noted that naming systems may not be as similar as they first appear—as for example in Fig. 1—for two reasons.

First, the similarity between Himba and Berinmo naming patterns for fully saturated stimuli (.61 inter-language agreement) does not extend to stimuli at low saturation (.27 inter-language naming agreement), for which the two languages' naming patterns are less similar than either is to English. Both Himba and English speakers use a large number of secondary terms to label desaturated stimuli, while Berinmo speakers readily extend basic terms to such stimuli. Reliance on the naming of maximally saturated stimuli may have led, in the past, to overestimation of the similarity of different languages' color term systems (Lucy, 1992; Lucy & Schweder, 1979; Saunders & van Brakel, 1997).

A second factor that contributes to the apparent similarity of the two figures is that, for simplicity, only the basic terms are shown with the range of each term defined by the name most frequently assigned to a particular chip. As was the case in Jameson and Alvarado (2003) and Roberson et al. (2000), Himba participants, in this study, were not restricted to the use of monolexemic terms or asked to make speeded decisions. In some areas, this resulted in particularly diffuse naming (as in the region that is covered by the term *purple* in English) and here the term assigned in Fig. 1 represents as little as 30% name agreement for several chips. Many observers either used a range of secondary terms or left a chip unnamed; a tendency much more prevalent amongst Himba speakers that Berinmo speakers reflected in the lower intra-language agreement between Himba speakers (.71) than Berinmo speakers (.83). Nevertheless, there are apparent similarities between the two naming systems.

There are several arguments that have been given for the origin of similar color categories that do not depend on linguistic relativity. One is that some adaptation of the visual system might result from learned characteristics of the environment because different observers experience different 'visual diets' (Mollon, 1982; Webster & Mollon, 1997). However, the different visual environments between our two languages with similar color terms would rule out the possibility. The main alternative arguments are rather those that derive from the overall similarity of color terms in the world's languages (Kay & Regier, 2003; Lindsey & Brown, 2002). These arguments essentially look for some common biological source for the origin of color categories.

One possibility is that differential photoxic effects of sunlight on the eye at different latitudes cause faster age-related deterioration of color discrimination in equatorial regions (Lindsey & Brown, 2002). Hence, similar terms are simply a result of damage to the eye from tropical conditions; thus, there is no need to develop color terms to separate categories in the blue/green regions. The present data are not suitable for testing that hypothesis as the Himba environment is of the type Lindsey and Brown could argue to produce retinal degeneration. In fact, if anything, it is the Berinmo data that would dispute their hypothesis as they live in lowland dense forest that sunlight does not easily penetrate. However, in a recent investigation, (Webster et al., 2002) showed systematic differences in the location of unique yellow judgments, in the absence of color-vision defects, within a population living in an equatorial region (India). These differences were attributed to cultural constraints (one population was composed of cloth factory workers who use particular shades of yellow thread) as the sunlight (UVB) exposure was equated for both groups (see also Hardy, Frederick,

Kay, & Werner, 2004). Moreover, all Himba participants in the present experiments (and the Berinmo participants tested by Roberson et al., 2000) had normal color vision, as measured by standard tests.

A more coherent argument for color term similarity comes from Berlin and Kay (1969) and their theory of categorization from universal prototypes. However, even given the general similarity, there are several reasons why the areas of color space taken by 5-term color languages need not be attributed to the driving force of universal prototypes. There are other restrictions on the possible color spaces. First, not all groupings potentially possible for an individual color term are logically coherent. Grouping by perceptual similarity (Roberson et al., 1999) precludes the formation of a category that includes, say, *red* and *yellow*, but excludes *orange*. The grouping-by-similarity 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 of languages with five basic terms is limited.

Second, and perhaps more important, as Jameson and Alvarado (2003) have argued, Berlin and Kay (1969) are likely correct in their argument for the initial divisions of color space, with an inevitable consequence of considerable similarity for subsequent color terms. If the first two terms are dark and light (dividing the apple along a lightness plane) and the third for reasons of biological importance (e.g., blood) is a hue term such as red, there are considerable constraints on subsequent divisions that would add the next couple of terms. It would be cognitively economical to have the two additional terms with centers maximally separated from each other and the other three terms. Thus, there is a likely similarity between color terms in the world's languages but this need not imply an underlying set of universal prototypes.

More critical to the theoretical debate, the present empirical data (Experiment 2) does not support theories of universal prototypes, nor does our recent work on shape categories (Roberson et al., 2002). The same conclusion has also been made for artifacts where distinctions in linguistic categorization apply for even the commonest objects (tables, Wierzbicka, 1992; containers, Malt et al., 1999). Such data led Malt and Sloman (in press, p. 5), to conclude that "naming must involve something more than, or different from, learning prototypes of universally perceived groupings". However, we do not wish to make the same case for linguistic relativity for artifacts that we do for color.

In studies of artifact categorization there is a genuine dissociation of naming from perceived similarity. Malt et al. (1999); Kronenfeld, Armstrong, and Wilmoth (1985) and Malt and Sloman (in press) found that, for artifact categories, judgments of the similarity of objects did not differ between speakers of languages who partitioned the objects into different name categories. It thus appears that perceptual categories (e.g., color) are differently susceptible to the influence of language than artifact categories. Malt and Johnson (1998) have argued that membership of artifact categories depends, at least to some extent, upon functional properties (e.g., the function to which an artifact was designed to be, or can be, put). But perceived similarity might depend, at least for the most part, on such perceptual properties as

color, shape, size, loudness, etc. However, perceptual categories are different. Similarity alone is not enough to ground perceptual categories since it provides no basis for deciding where to place boundaries (see Roberson et al., 1999 or Dummett, 1975 for a discussion of the Sorites' paradox).

Perceptual continua such as color may thus be a special case for categorization with the consequence that the influence of culture (and language as the instrument of culture) may be strongest just for those 'fuzzy' sets for which there are not obvious discontinuities in nature. Indeed, our recent developmental studies show that Himba children behave like English children in making color similarity judgments when both know no color names (Roberson et al., in press). Initially, both judge color similarity on perceptual grounds. Thereafter, the origins of the color categories in different societies might be constrained by different cultural or environmental needs (Nisbett, Peng, Choi, & Norenzayan, 2001; Sera et al., 2002; Wierzbicka, 1990, 1992), but this question is beyond the scope of the present study. Whatever the origin of the observed differences between the color terminologies of different societies, linguistic categorization, in adults, appears isomorphic with cognitive representation. Perceptual space appears to be distorted at the boundaries of color categories, so that, even when two languages have the same number of terms and those terms cluster around similar points in perceptual space, speakers of those languages show significant differences in their cognitive organization of color space. Thus, when considering whether two sets of categories are effectively equivalent, the position of the category boundaries should be considered of, at least, equal importance with the category centers.

Acknowledgment

Appendix A. Portable Munsell light-box used in Experiments 1 and 3

A.1. Dimensions

External Dimensions: With handle to top – 390 mm high × 465 mm wide × 200 mm deep.
Weight: 17 kg/40 lb fully packed.
Contents not built in: spare battery, mains power unit, 2 × spare tubes, 2 × box side panels.

A.2. Specification

Case: lockable IP65 sealed equipment case, with detachable externally fitted solar panel and temperature sensor.

Fitted with immersion-proof breather for pressure equalization in flight and water proofing when case locked.

Light head: flipper locks and hinged light head housing 2×300 mm, 8 W illuminant C tubes and $4 \times$ tube starters (working right hand ones, $2 \times$ spares to left of head).

Power section: to the right of the base unit is housed one sealed lead acid battery and high voltage inverter to power lamps and inline fuse holder only.

Battery: high specification cyclic rechargeable 12 V/6.5 Ah with 15 A protection fuse to upper case. Re-sealable vents for over-temperature and charging misuse.

Inverter protected by a 30 A fuse and bonded to case for safety. Delivers approximately 240 V AC to power lamp head.

Control section: Control panel to right of hinged lamp head in top of case. Components behind panel: lamp ballast choke, $4 \times$ control switches, Voltage meter, Solar shunt for panel voltage control, connection block and all interconnecting wires.

Case also contains temperature and humidity monitors mounted inside upper case, with internal and external sensors.

Appendix B. Percentage of individuals giving the same label to each stimulus in the set used in Experiment 3

Stimulus	Called "blue" by English speakers (%)	Called "nol" by Berinmo speakers (%)	Called "burou" by Himba speakers (%)
7.5B	100	100	96
5B	100	100	94
2.5B	100	100	96
10BG	87	100	94
7.5BG	51	100	94
5BG	09	100	92
2.5BG	0	100	94
10G	0	100	96
7.5G	0	100	94
	Called "green" by English speakers (%)	Called "nol" by Berinmo speakers (%)	Called "burou" by Himba speakers (%)
5Y	67	0	0
7.5Y	82	0	0
10Y	94	0	0
2.5GY	100	06	02
5GY	100	68	04
7.5GY	100	92	18
10GY	100	96	94
2.5G	100	100	96
5G	100	100	94

Appendix C. Stimulus pairs used in Experiment 3b

blue–green within-category *blue–green* cross-category:

5B–2.5B 2.5BG–7.5BG
5B–10BG 10BG–5BG
5BG–2.5BG 7.5BG–5BG
5BG–10G 7.5BG–2.5BG

dumbu–burou within-category *dumbu–burou* cross-category:

7.5Y–10Y 5GY–7.5GY
7.5Y–2.5GY 2.5GY–7.5GY*
2.5GY–5GY* 7.5GY–10GY*
10Y–5GY* 7.5GY–2.5G

* The status of these pairs is reversed for the Berinmo categories *nol–wor*.

Appendix D. Munsell designations of stimuli used for category learning in Experiment 3c

green–blue set *green1–green2* set:

2.5B 3/8, 4/10, 5/8, 6/6, 7/8 2.5G 3/8, 4/6, 5/10, 6/8, 7/8
10BG 3/8, 4/6, 5/10, 6/8, 7/8 5G 3/8, 4/6, 5/10, 6/8, 7/10
5BG 3/8, 4/6, 5/10, 6/10, 7/8 10G 3/8, 4/8, 5/6, 6/8, 7/8
2.5BG 3/8, 4/6, 5/10, 6/8, 7/8 2.5BG 3/8, 4/6, 5/10, 6/8, 7/10

yellow–green set *dumbu–burou* set:

2.5Y 8/8, 8/10, 8/12 7.5Y6/6, 6/8, 6/10, 7/10
5Y 7/10, 7/12, 8/8, 8/10, 8/12 2.5GY 6/6, 6/8
7.5Y 8.5/10, 8.5/12, 7.5GY 5/6, 6/6,6/10, 6/12
2.5GY 7/10, 7/12, 8/8, 8/10, 8/12 5GY 8/8, 8/10, 8/12
7.5GY 8/8, 8/10 2.5/G 5/10, 6/10, 7/8
10GY 6/8, 6/10, 7/10 5G 5/6, 6/6, 6/8, 6/10

References

Berlin, B., & Kay, P. (1969). *Basic color terms: Their universality and evolution*. Berkeley: University of California Press.
Booysen, J. M. (1987). *Otjiherero*. Windoek: Gamsberg.
Bornstein, M. H. (1987). Perceptual categories in vision and audition. In S. Harnad (Ed.), *Categorical perception: The groundwork of cognition*. Cambridge: Cambridge University Press.

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.

Boroditsky, L. (in press). Linguistic relativity. *Encyclopedia of Cognitive Science*. MacMillan Press (to appear).

Boroditsky, L. (2001). Does language shape thought. Mandarin and English speakers' conceptions of time. *Cognitive Psychology, 43*, 1–22.

Bowerman, M., & Choi, S. (2001). Shaping meanings for language: Universal and Language specific in the acquisition of spatial semantic categories. In M. Bowerman & S. Levinson (Eds.), *Language acquisition and conceptual development*. Cambridge: Cambridge University Press.

Brown, R. (1976). In memorial tribute to Eric Lenneberg. *Cognition, 4*, 125–153.

Crandall, D. P. (2000). *The place of stunted ironwood trees: A year in the lives of the cattle-herding Himba of Namibia*. New York: Continuum—Academi.

Clarke, M., Lossoff, A., McCracken, M., & Rood, D. (1984). Linguistic relativity and sex/gender studies: Epistemological and methodological considerations. *Language Learning, 34*, 47–67.

Clarke, M., Lossoff, A., McCracken, M., & Still, J. (1981). Gender perception in Arabic and English. *Language Learning, 31*, 159–169.

Davidoff, J., Davies, I. & Roberson, D. (1999) Colour categories of a stone-age tribe. *Nature, 398*, 203–204; *402*, 604.

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

Dummett, M. (1975). Wang's Paradox. *Synthese, 30*, 301–324.

Fletcher, R. (1980). *The City University Colour Vision Test*. Windsor, Berks: Keeler Ltd.

Gennari, S. P., Sloman, S., Malt, B., & Fitch, W. T. (2000). Motion events in language and cognition. *Cognition, 83*, 49–79.

Goldstone, R. (1994). Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology: General, 123*, 178–200.

Goldstone, R. L. (1998). Perceptual learning. *Annual Review of Psychology, 49*, 585–612.

Guest, S., & Van Laar, D. (2002). The effect of name category and discriminability on the search characteristics of color sets. *Perception, 31*, 445–461.

Gumperz, J. J., & Levinson, S. C. (1997). Rethinking linguistic relativity. *Current Anthropology, 32*, 613–623.

Hardy, J. L., Frederick, C. M., Kay, P., & Werner, J. S. (2004). Color naming, lens aging and grue: What the optics of the aging eye can teach us about color language. *Psychological Science, 15*, 289–290.

Harnad, S. (1987). Psychophysical and cognitive aspects of categorical perception: A critical overview. In S. Harnad (Ed.), *Categorical perception: The groundwork of cognition*. Cambridge: Cambridge University Press.

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

Ishihara, S. (1992). *Ishihara's tests for colour-blindness*. Tokyo: Kanehara & Co. Ltd.

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

Jordan, G., & Mollon, J. D. (1997). Adaptation of color vision to sunlight. *Nature, 386*, 135–136.

Karmiloff-Smith, A. (1979). *A functional approach to child language: A study of determiners and reference*. Cambridge, UK: Cambridge University Press.

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

Kay, P., & Kempton, W. (1984). What is the Sapir–Whorf hypothesis?. *American Anthropologist, 86*, 65–78.

Kay, P., & McDaniel, C. K. (1978). The linguistic significance of the meanings of basic color terms. *Language, 54*, 610–646.

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

Kronenfeld, D. B., Armstrong, J. D., & Wilmoth, S. (1985). Exploring the internal structure of linguistic categories: An extensionist semantic view. In J. W. D. Dougherty (Ed.), *Directions in cognitive anthropology*. Urbana: University of Illinois Press.

Legendre, P. (2000). Comparison of permutation methods for the partial correlation and partial Mantel tests. *Journal of Statistics and Computer Simulation, 67*, 37–73.

Levinson, S. C. (1996). Language and space. *Annual review of Anthropology, 25*, 353–382.

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

Lucy, J. A. (1992). *Language and diversity of thought: A reformulation of the linguistic relativity hypothesis.* Cambridge: Cambridge University Press.

Lucy, J. A., & Schweder, R. A. (1979). Whorf and his critics: Linguistic and nonlinguistic influences on color memory. *American Anthropologist, 81*, 581–605.

MacLaury, R. E. (1987). *Color and cognition in Mesoamerica: Constructing categories as vantages.* Austin: University of Texas Press.

Malt, B. C., & Johnson, E. C. (1998). Artifact category membership and the intentional–historical theory. *Cognition, 66*, 79–85.

Malt, B. C. & Sloman, S.A. (2003). Linguistic diversity and object naming by non-native speakers of English. Bilingualism: *Language and Cognition, 6*, 46–67.

Malt, B. C., Sloman, S. A., Gennari, S., Shi, M., & Wang, Y. (1999). Knowing versus naming: Similarity and the linguistic categorization of artefacts. *Journal of Memory and Language, 40*, 230–262.

Mantel, R. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Research, 27*, 209–220.

Mollon, J. D. (1982). Color-vision. *Annual Review of Psychology, 33*, 41–85.

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.

Namibian Government statistics. (2004). Available from: <www.grnnet.gov.na/>.

Nisbett, R. E., Peng, K. P., Choi, I., & Norenzayan, A. (2001). Culture and systems of thought: Holistic versus analytic cognition. *Psychological Review, 108*, 291–310.

Özgen, E. (2004). Language, learning, and color perception. *Current directions in Psychological Science, 13*, 95–98.

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

Pérez-Pereira, M. (1991). The acquisition of gender: What Spanish children tell us. *Journal of Child Language, 18*, 571–590.

Pilling, M., Wiggett, A., Özgen, E., & Davies, I. R. L. (2003). Is color categorical perception really perceptual?. *Memory & Cognition, 31*, 538–551.

Ray, V. F. (1952). Techniques and problems in the study of human color perception. *Southwestern Journal of Anthropology, 8*, 251–259.

Roberson, D., & Davidoff, J. (2000). The categorical perception of colours and facial expressions: The effect of verbal interference. *Memory & Cognition, 28*, 977–986.

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

Roberson, D., Davidoff, J., & Shapiro, L. (2002). Squaring the circle: The cultural relativity of good shape. *Journal of Cognition and Culture, 2*, 29–53.

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

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

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

Saunders, B., & van Brakel, J. (Eds.). (2002). *Theories, technologies, instrumentalities of color: Anthropological and historical perspectives.* Lanham, MD: University Press of America.

Sera, M. D., Berge, C., & del Castillo Pintado, J. (1994). Grammatical and conceptual forces in the attribution of gender by English and Spanish speakers. *Cognitive Development, 9*, 261–292.

Sera, M. D., Elieff, C., Forbes, J., Burch, M. C., Rodriguez, W., & Dubois, D. P. (2002). When language affects cognition and when it does not: An analysis of grammatical gender and classification. *Journal of Experimental Psychology: General, 131*, 377–397.

Stefflre, V., Castillo Vales, V., & Morley, L. (1966). Language and cognition in Yucatan: A cross-cultural replication. *Journal of Personality and Social Psychology, 4*, 112–115.

Webster, M. A., & Mollon, J. D. (1997). Adaptation and the color statistics of natural images. *Vision Research, 37*, 3283–3298.

Webster, M. A., Webster, S. M., Bharadwaj, S., Verma, R., Jaikumar, J., Madan, G., & Vaithilingham, E. (2002). Variations in normal color vision III. Unique hues in Indian and United States observers. *Journal of the Optical Society of America, 19*, 1951–1962.

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

Wierzbicka, A. (1992). *Semantics, culture and cognition.* Oxford: Oxford University Press.

Wyszecki, G., & Stiles, W. S. (1982). *Color science.* New York: Wiley & Sons.