Categorical perception of animal patterns

Julie Goldstein and Jules Davidoff*

Department of Psychology, Goldsmiths, University of London, London, UK

As part of the more general issue of whether culture can affect perception, the present paper addresses the Whorfian question of whether the language available to describe perceptual experience can influence the experience itself. It investigated the effect of vocabulary on perceptual classification by the study of a remote culture (Himba) which possesses a poor colour vocabulary but a rich vocabulary of animal pattern terms. Thus, the present study examined Categorical Perception (CP) with a type of visual stimulus not previously used to assess the effect of labels on perceptual judgments. For the animal patterns, the Whorfian view predicted that it would only be the Himba who showed superiority for cross-category decisions as only they have the appropriate labels. The Whorfian view was upheld and confirmed previous findings that linked perceptual differences to labelling differences.

As part of the more general issue of whether culture can affect perception (Nisbett & Miyamoto, 2005), the present paper addresses the Whorfian question of whether the language available to describe perceptual experience can influence the experience itself (Carroll, 1956). There has been a considerable revival of interest in the Whorfian hypothesis (Lucy, 1992; Davidoff, Davies, & Roberson, 1999; Boroditsky, 2001; Özgen & Davies, 2002; Saunders & van Brakel, 2002). The revival has provided evidence for the Whorfian view from systematic investigations of the relationship between language and thought. Differences between languages in their grammatical structure and vocabulary have been associated with perceptual differences of the same experience in the following domains: grammatical gender (Sera, Berge, & Del Castillo Pintado, 1994; Sera *et al.*, 2002, Boroditsky, Schmidt, & Phillips, 2003); colour (Kay & Kempton, 1984; Davidoff *et al.*, 1999; Roberson, Davidoff, Davies, & Shapiro, 2004); material and shape classification (Lucy, 1992); spatial relations (Levinson, 1996; Bowerman & Choi, 2001), number systems (Gumpertz & Levinson, 1996, Gordon, 2004), artifact categories (Malt & Johnson, 1998); modes of motion (Gennari, Sloman, Malt, & Fitch, 2000); time (Boroditsky, 2001) and shape (Roberson, Davidoff, & Shapiro, 2002). However, other recent studies have argued against the influence of linguistic differences on mathematical abilities (Gelman & Butterworth, 2005), colour (Kay & Regier, 2003), and perceptual classification, both at the level of

* Correspondence should be addressed to Professor Jules Davidoff, Department of Psychology, Goldsmiths, University of London, Lewisham Way, London SE14 6NW, UK (e-mail: j.davidoff@gold.ac.uk).

terminology (Munnich & Landau, 2003; Malt, Sloman, Gennari, Shi, & Wang, 1999) and grammatical structure (Karmiloff-Smith, 1979; Perez-Pereira, 1991). The present study seeks to provide further evidence in favour of an effect of vocabulary on perceptual classification by the study of a remote culture which has a rich vocabulary of pattern terms.

The study was carried out with the Himba who are a semi-nomadic population of animal herders estimated between 20,000 and 50,000 (Namibian Government statistics, 2004) whose territory spreads over an area of approximately 25,000 square miles, in northern Namibia and Angola. They inhabit an arid region; their visual diet is of open desert, scrubland, and mountain. The Himba speak a dialect of Herero but, unlike that group, have retained a strong and distinctive traditional cultural identity that brings little contact with other cultures. See Crandall (2000) for an account of the Himba as a distinct, cohesive cultural, and linguistic group.

Eckl (2000), and Turton (1980) for an Ethiopian group of cattle herders with similar life-style to that of the Himba, outlines cultural interests and reasons why colour and patterns are used to recognize individual animals. Therefore, it is not surprising that the Herero and Himba have many names for cattle patterns (Eckl, 2000). The present study will assume that each of these names refers to a category of patterns and will argue that these categories should obey the principles of Categorical Perception (CP). Harnad (1987) provides a comprehensive discussion of CP across a range of auditory and visual categories. He argued that, with CP, a physical continuum is perceived as qualitatively discontinuous. Items from different categories appear more different than items from the same category despite being equated for physical difference (e.g. by numbers of just noticeable differences). CP is thereby revealed by faster and more accurate discriminations between stimuli that cross a category boundary than between two stimuli that are both good exemplars of the same category (Bornstein, 1987). Goldstone (1994) further argued that this between-category expansion might also be accompanied by a within-category compression where items within a category become to look more similar. As a technique, CP has been used to consider debates in face (Etcoff & Magee, 1992) and colour categorization (Roberson, Davidoff, & Braisby, 1999).

In cross-cultural studies, CP has been used to show precise differences based on colour terms in the speaker's language (Roberson, Davidoff, Davies, & Shapiro, 2005). However, there has been considerable debate as to whether the origin of colour CP is linguistic (Roberson *et al.*, 2004, 2005) or at least is partially based on universal neurophysiological constraints in colour vision (Bornstein, 1975; Franklin & Davies, 2004). No other visual type of stimulus has been used in cross-cultural investigations of CP. The present paper by extending the CP methodology to animal patterns does so clearly to categorization that is not innately given. It will use the two-alternative forced-choice (2AFC) methodology as in the previous studies (Roberson, Davies, & Davidoff, 2000) to examine whether a target is better recognized when contrasted with a foil from a different category when compared with when the foil is from the same category. In summary, the aim of this study was to provide a different type of stimulus to confirm previous findings with colour that linked perceptual and cognitive differences to labelling differences.

Production of the animal continua

Stage 1: Preliminary naming

Two individuals from different Himba villages (man, aged 45 years; woman, aged 35 years) helped in the picture taking of animals at two different waterholes. With their advice, 128 pictures were taken with a digital camera; these were transferred to a laptop

pc and shown to 10 Himba men (aged 23–45 years), from two separate groups. Like all the subsequent participants, they were rural inhabitants of Northern parts of Namibia (Kaokoveld); mainly herders taking their animals to water sources, often having to travel quite far by foot. They were monolingual and paid in kind. All participants were screened for colour vision abnormalities with the City Colour Vision Test (Fletcher, 1980). Every participant was tested individually in a darkened vehicle. The experimenter showed one picture at a time on a computer, and the translator asked the participant to give a name per picture and recorded the answers in a notebook.

Stage 2: Obtaining the morphed stimuli

Eighteen pictures were selected on the basis of results from naming Stage 1; for these pictures, there was 90% naming agreement. The animals were isolated from their context/background and were manipulated in Adobe Photoshop and given a light yellow background, approximating the colour of the earth/sand of the Himba territory. Next, pictures were paired by the experimenter and manipulated in the Meditor morphing package to produce a continuum. Every continuum was based on two pictures (end-points) from the same species (cow, goat, or sheep) and resembled each other to increase the likelihood of a 'natural' continuum after morphing (see Figure 1 for end-points used in the 2AFC Test). However, to ensure that shape of the animal was identical for both end-points, the same silhouette was used for both end-points. The pattern was lifted from one end-point in Photoshop and transferred to the outline of the other end-point. A continuum consisted of 21 pictures in total as there was a 5% difference between each morphed picture going from one end-point to the other. Nine continua were created following this procedure, resulting in 189 pictures (21 pictures × 9 continua) that were printed and prepared for use in Stage 3.

Stage 3: Naming the morphed continua

For the second phase of naming, laminated pictures were placed on a table either in a sequential or in a random order. The sequential order (1–21 in Figure 2) was simpler to administer but as this could have led to perseveration in naming, we added a random order condition. Fifteen different Himba from those in Stage 1 (12 men and 3 women, varying in age from 17 to 45 years) participated in the sequential naming, and 15 different Himba from those in Stage 1 and those who did the sequential order (12 men and 3 women varying in age from 19 to 50 years) participated in the naming in random order. Every Himba participant was asked to give the name of the animal that the experimenter pointed at, or was asked to name every picture individually (depending on the condition). In the sequential phase, four blocks were needed to assess 189 pictures, as only 50 pictures fitted the table. For individual picture naming, the Himba participant named 189 pictures placed on the table in random order. All answers were recorded on an answer sheet.

There was some disagreement about the names given in phase 2 to those in phase 1. Our interpreter advised us that this was to be expected as the Himba participants came from a large area and that there was some naming variation between different groups.

Stage 4: Deciding on the continua

The fourth stage in the production of the animal continua was to ensure that they would be suitable for use in a 2AFC experiment. In one respect, all continua were suitable. In all continua (see, for example, Figure 2a–e), there was a sharp drop where the name given to one end-point changes to the name given to the other end-point. As these intermediary

(a) Continuum 1

(b) Continuum 2

(c) Continuum 3

(d) Continuum 4

(e) Continuum 5
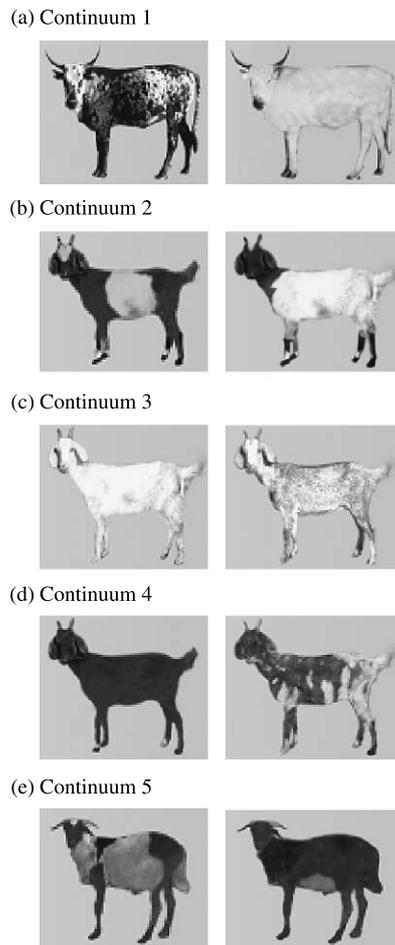
Figure 1. (a–e) End-point pictures of continua used for 2AFC test.

points represent 'imaginary' animals, one might ask why there is not a gradual change from one name to the next rather than a sudden change. The most likely answer is that we have chosen prototypical animals for the end-points. Animals with the same name would bear a strong family resemblance to the prototype, and the morphing procedure would move, in a few images, from pictures that would easily be named as the end-point to ones that would not be so named. Of course, there was an area on every continuum that some of the Himba were not prepared to give either end-point names and gave other names. However, for a different reason, not all nine continua could be used. Four of the nine continua consisted of boundaries (name changes) that were very close to either one of the end-points. It was therefore not possible to create a 2AFC procedure for these continua as the test requires a minimum of a boundary location at five pictures away from the end-points.

The remaining five continua were suitable for use in the 2AFC test. The agreement on end-point names and our interpreter's definition of names were as follows: Continuum 1, Cow (Ongange–Omvahe), 90% consensus for Ongange (white spots on body and face) and 56.66% for Omvahe (cream colour); Continuum 2, Goat (Orupera–Ongwiti) 53.33% agreement for Orupera (brown body with white circling rim

(a) Continuum 1

(b) Continuum 2

(c) Continuum 3

(d) Continuum 4
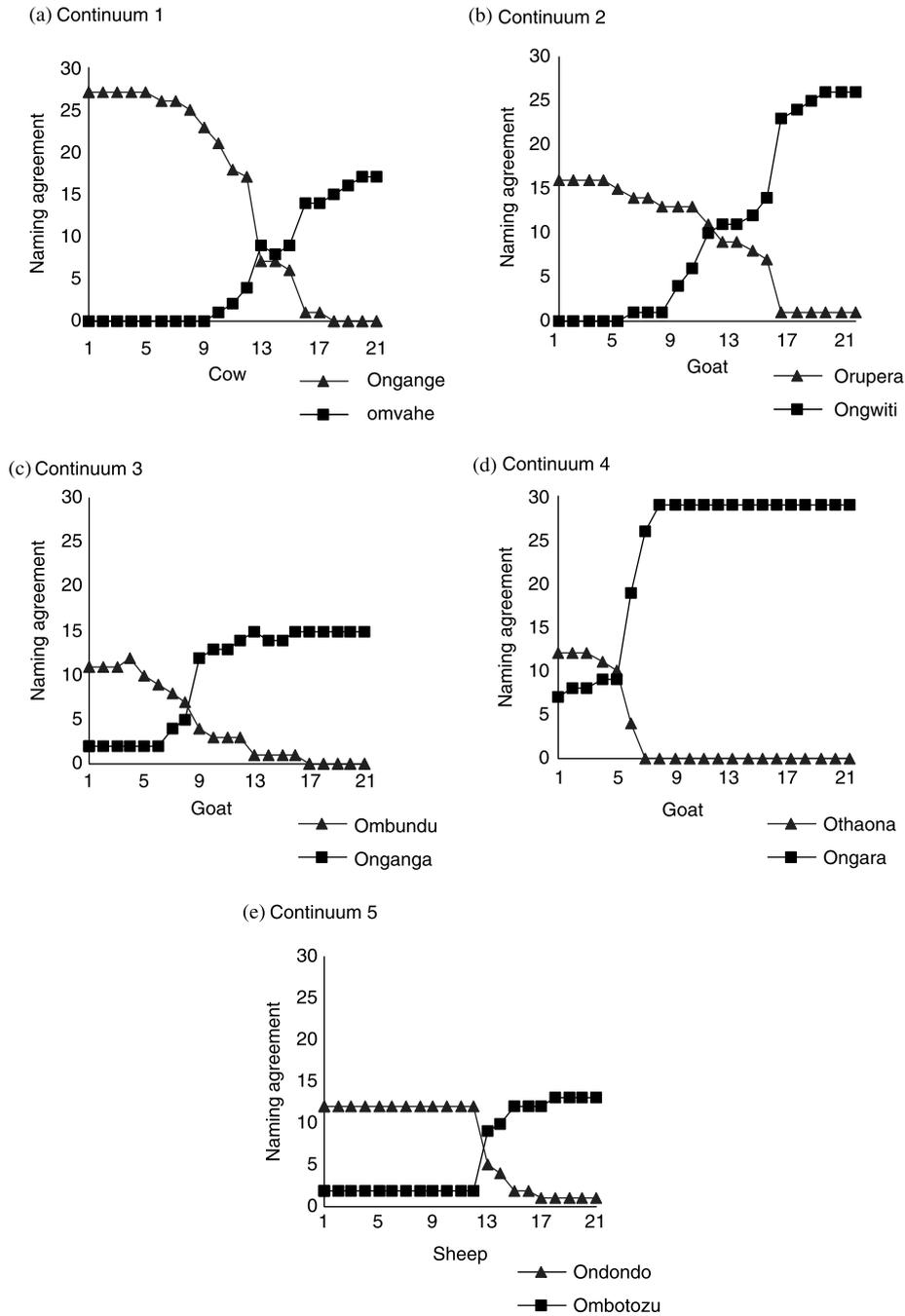
(e) Continuum 5

Figure 2. (a–e) Naming agreement from 30 Himba participants on 5 continua for animal patterns. Each continuum consisted of 21 pictures (19 equally spaced morphed images from the two end-points). Continuum 1 (a) was for cow patterns, Continua 2 (b), 3 (c), and 4 (d) were goat patterns and Continuum 5 (e) was for sheep patterns. The two curves for each continuum represent the number of times the end-point names were given to each picture.

around the body) and 86.66% for Ongwiti (white body and brown or black head and neck); Continuum 3, Goat (Ombundu–Onganga), 36.66% consensus for Ombundu (light red body) and 50% for Onganga (tiny dots all over a red body); Continuum 4, Goat (Othaona–Ongara), 40% consensus for Othaona (golden-brown colour) and 96.66% for Ongara (black background with golden-beige patches); Continuum 5, Sheep (Ondondo–Ombotozu), 40% consensus for Ondondo (white with black around the body and ears) and 43.33% for Ombotozu (brown/black with white on the stomach). To ensure that the two naming techniques (sequential or random) from Stage 3 did not cause a difference in naming, t tests were performed on the naming scores from each technique comparing each of the 21 positions for each continuum. There were no significant differences. All t values were < 1 except Continuum 3 (Onganga, $t(20) = 1.2$, $p = .25$) and Continuum 4 (Ongara, $t(20) = 1.48$, $p = .15$).

The boundary at which one name changed to the other was assessed by eye from the name agreement at all 21 positions (see Figure 2a–e). For Continuum 1, the boundary was between pictures 12 and 13. For Continuum 2, the boundary was between pictures 15 and 16. For Continuum 3, the boundary was between pictures 6 and 7. For Continuum 4, the boundary was between pictures 5 and 6, and for Continuum 5, the boundary was between pictures 12 and 13.

2AFC TEST

Method

Participants
Thirty different Himba (17 women and 13 men, varying in age from approximately 18 to 45 years) and 30 English (17 women and 13 men varying in age from approximately 20 to 43 years) took part in the 2AFC test. The English participants were students at Goldsmiths' College and participated under course requirements. As for the production of the animal continua, the Himba were monolingual, rural inhabitants of Northern parts of Namibia (Kaokoveld), tested in those areas and paid in kind. All participants were screened for colour vision abnormalities with the City University Colour Vision Test (Fletcher, 1980).

Materials
Pictures for the between-category and within-category items were selected from the five continua (see Stage 4 of preceding section). There was always 10% morph length between two pictures in each pair. Figure 3a–e shows the section of each continuum used to select pictures for the 2AFC range. It also shows the actual pictures used for each type of trial. Note that though the same labels (a–j) are used in Figure 3 to denote the section of each continuum used in the 2AFC, these came (see Stage 4 above) from different positions on the 21-step morphed continua.

Procedure
Practice
The task was explained to the Himba via an interpreter. He told the participant that he would show a picture of an animal and that they had to remember what they had seen.

(a) Continuum 1

a   b   c   d   e   f   g   h   i   j

{boundary}

(b) Continuum 2

a   b   c   d   e   f   g   h   i   j

{boundary}

(c) Continuum 3

a   b   c   d   e   f   g   h   i   j

{boundary}

(d) Continuum 4

a   b   c   d   e   f   g   h   i   j

{boundary}

(e) Continuum 5

a   b   c   d   e   f   g   h   i   j
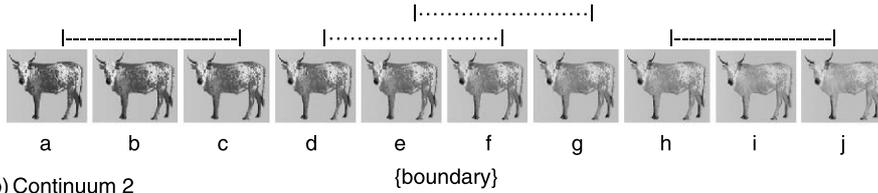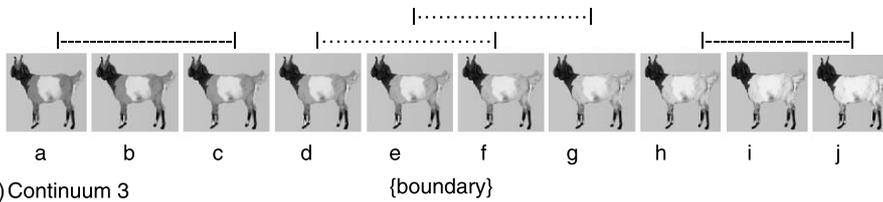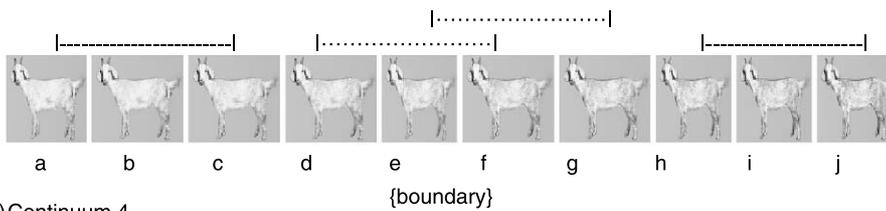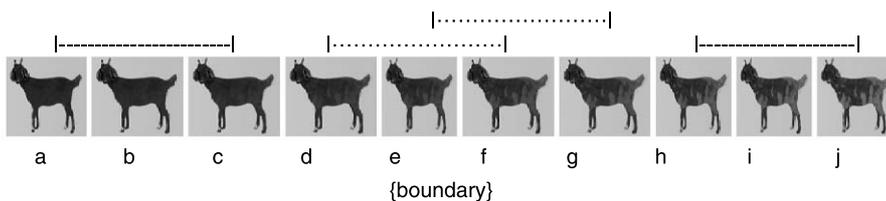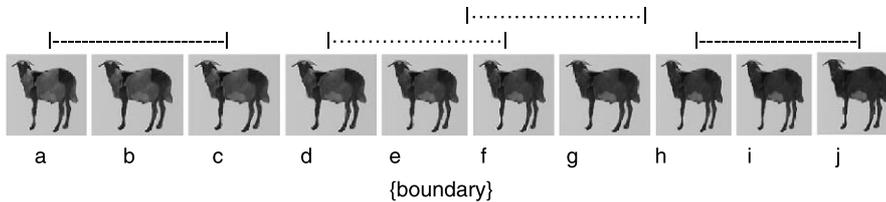
{boundary}

Figure 3. (a–e) Subsets of the five continua containing the pictures used in the 2AFC test. The boundary for each continuum was between e and f. Within-category pairs were pictures a–c and h–j (|- - -|). Between-category pairs were d–f and e–g (| . . . |).

After the first picture, there would be immediately two pictures one of which was identical to the first. The participant was asked to point to the one that was identical. Practice trials (maximum of 10) with feedback were given using animals not included in the test until there were three consecutively correct answers. Exactly the same procedure was adopted for the English participants except that the stimuli were shown on a computer (programmed in E-prime) with a 5-second interval between target and

2AFC. The participants were asked to respond by pressing key 'v' for the left picture and 'n' for the right picture when presented with the two alternatives.

Test

Procedure at test was similar to that at practice. The experiment was carried out manually in Namibia in a shaded area and answers were recorded on a scoring sheet. The test consisted of 40 2AFC trials per continuum (20 cross-category and 20 within-category trials), summing to 200 trials per participant. Half of the trials for each continuum had the correct answer on the left and half the correct answer on the right. A different random order of the 200 trials was administered to each participant with trials for all the continua intermingled. To ensure smooth running of the experiment, for the Himba, the trial cards were assembled prior to the start of the experiment.

Naming

At the end of the test, Himba participants were shown all pictures of the test one at a time and asked to name them.

Results

Naming

Naming agreement was assessed for the eight pictures used for each continuum. Each picture was classified as in agreement if the same name was given as in the production of the animal continua (see Stage 4 of previous section). For each continuum, some Himba participants were in complete agreement but others showed disagreement on one name or more. The numbers of Himba with naming agreement for the continua were as follows: 27 for Continuum 1; 22 for Continuum 2; 10 for Continuum 3 (and therefore 20 participants in naming disagreement); 29 for Continuum 4; and 19 for Continuum 5 (and 11 participants in naming disagreement).

2AFC

For the initial analysis, all Himba and English participants were included to allow all continua to be included in the same analysis. A 2 (Group: Himba vs. English) $\times$ 2(Trial Types: Between-Category vs. Within-Category) $\times$ 5 (Continuum: Continua 1 vs. 2 vs. 3 vs. 4 vs. 5) ANOVA, with repeated measures over the second and third factors was carried out on the accuracy data. All main effects were significant. The Himba were more accurate than the English ($F(1, 58) = 7.84$, $p = .007$), Between-Category trials were more accurate than Within-Category Trials ($F(1, 58) = 11.36$, $p = .001$) and Continua differed in their recognition accuracy ($F(4, 232) = 11.37$, $p < .001$, with highest performance on Continua 1 and 4 and lowest scores on Continua 2 and 5).

In the context of these main effects, the analysis showed three significant two-way interactions. Most important, there was a significant Group $\times$ Trial Types interaction ($F(1, 58) = 32.62$, $p < .001$), revealing on *post hoc* analysis (Bonferroni correction, $p < .013$), higher scores for the Himba on Between-Category trials than Within-Category trials, $t(29) = 5.34$, $p < .001$, whereas English performance did not reach reliable levels between Trial Types, $t(29) = 2.22$, $p = .034$ which, in any case was in the direction of greater accuracy for Within-Category trials. The Himba were more accurate than the English on Between-Category trials, $t(29) = 5.62$, $p < .001$, but the Groups did not differ on Within-Category trials, $t(29) < 1$.

There were also reliable effects that involved Continuum. The Category × Continuum interaction gave $F(4, 232) = 10.2$, $p < .001$ and there was also a weakly significant Group × Continuum interaction, $F(4, 232) = 2.43$, $p = .048$. In light of these effects that involved differences between continua and also because Himba were included in the analysis who did not agree on the names for the animal patterns, further analyses were carried out on individual continua. Analyses were carried out on the accuracy data for Continua 1, 2, and 4 on only the participants who were in naming agreement. However, in Continua 3 and 5, there was substantially more naming disagreement than for other continua, so Himba naming disagreement groups were also considered in the data analyses of these continua.

Continuum 1

A 2 (Group: Himba vs. English) × 2 (Trial Types: Between-Category vs. Within-Category) ANOVA, with repeated measures on the second factor was carried out on the accuracy data. It showed a significant interaction ($F(1, 55) = 11.18$, $p = .001$) accompanied by main effects of Trial Type ($F(1, 55) = 21.87$, $p < .001$) with better performance for Between-Category Trials, and Group ($F(1, 55) = 10.28$, $p = .002$) with better performance for the Himba (see Figure 4a). Analysis of the interaction revealed higher scores for the Himba on Between-Category trials ($t(55) = 4.6$, $p < .001$) but not on Within-Category trials ($t(55) = 1.39$, $p > .1$). However, the Himba ($t(26) = 5.58$, $p < .001$) have a Between-Category advantage over Within-Category trials whereas the English do not ($t < 1$). The lowest mean accuracy ($M = 12.80$, $SD = 2.34$) was for the English on Within-Category trials, and this score was above chance ($t(29) = 6.55$, $p < .001$).

Continuum 2

A 2 (Group: Himba vs. English) × 2 (Trial Types: Between-Category vs. Within-Category Trials) ANOVA, with repeated measures on the second factor, was carried out on the accuracy data. It only showed a significant interaction ($F(1, 50) = 13.41$, $p = .001$). Both Main effects (Trial Type and Group) gave $F < 1$ (see Figure 4b). Analysis of the interaction revealed higher scores for the Himba on Between-Category trials ($t(50) = 2.57$, $p = .013$), but not on Within-Category trials ($t(50) = 1.37$, $p > .1$). However, unlike for Continuum 1, the Himba only showed a trend for superior performance on Between-Category trials ($t(21) = 1.74$, $p = .097$) whereas the English ($t(29) = 3.17$, $p = .001$) were more accurate on Within-Category trials. A further analysis showed that much higher accuracy was obtained for English participants on one of the Within-Category pairs ($t(29) = 3.88$, $p = .001$). It was for the Within-Category pair h–j in Figure 3b. As one may see, the contrast in picture j is low and could be remembered as having a much larger white area than the other. The lowest mean accuracy ($M = 10.80$, $SD = 2.54$) was for the English on Between-Category trials; this score was not above chance ($t(29) = 1.73$, $p > .05$). The next lowest mean accuracy ($M = 11.68$, $SD = 2.38$) was for the Himba on Within-Category trials; this score was above chance ($t(21) = 3.32$, $p = .003$).

Continuum 3

A 3 (Group: Himba Naming Agreement (HNA) vs. Himba Naming Disagreement (HND) vs. English) × 2 (Trial Types: Between-Category vs. Within-Category) ANOVA, with repeated measures on the second factor was carried out on the accuracy data. The main
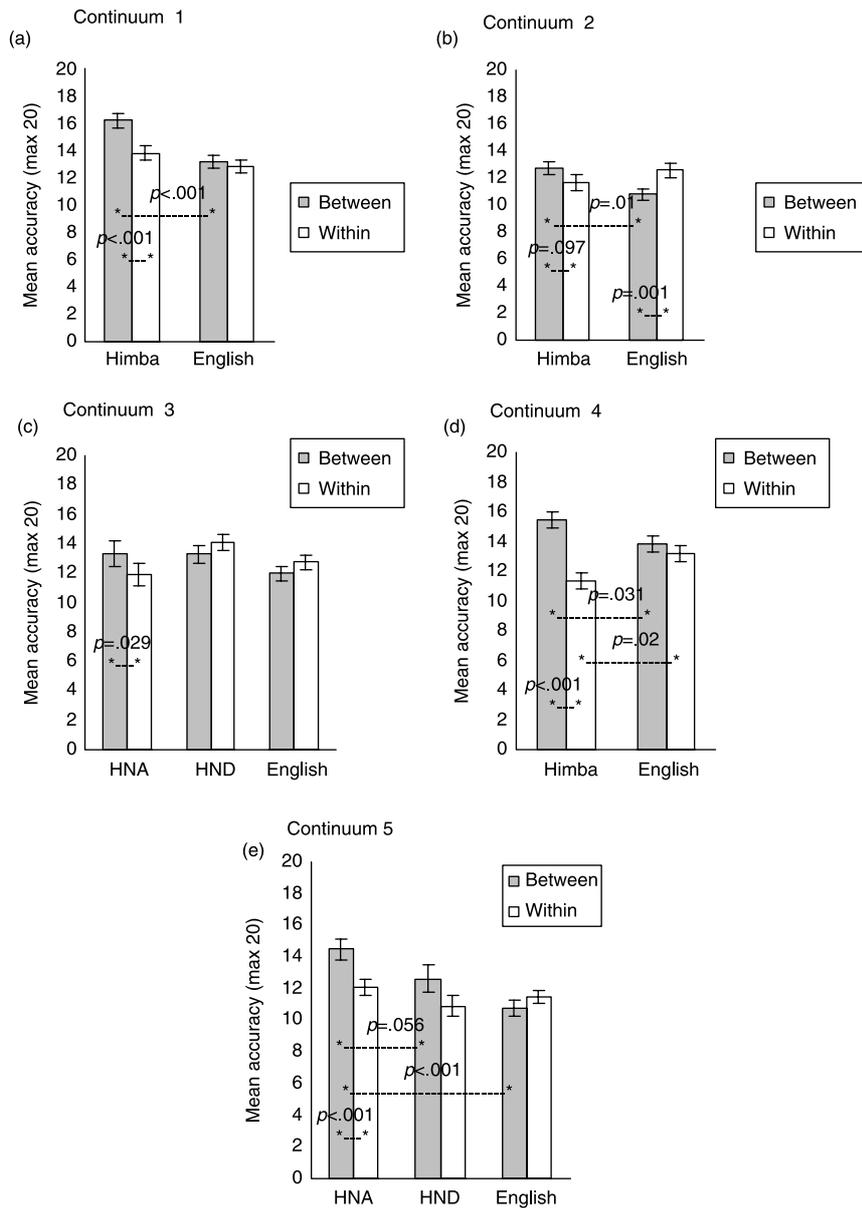
Figure 4. (a–e) Mean accuracy scores (max = 20) and standard errors on five continua (a–e) on Between-Category and Within-Category trials in the 2AFC test. Scores are given separately for English and Himba participants. Himba participants for Continua 3 and 5 are divided into those with naming agreement (HNA) and those in disagreement (HND). All comparisons giving $p < .1$ are indicated for each continuum.

effect of Trial Type gave $F < 1$, no effect of Group ($F(2, 57) = 2.30, p > .1$) was found, and the interaction showed a trend ($F(2, 57) = 2.49, p = .092$) (see Figure 4c). This trend was further explored in *post hoc* tests. When comparing trial types within groups, the only significant value found was for HNA with a between-category

advantage ($t(9) = 2.58$, $p = .029$). A similar comparison for HND gave $t(19) = −1.67$, $p > .1$, and for English gave $t(29) = 1.18$, $p > .2$. When comparing groups across a single trial-type condition, the groups failed to show significantly different performances from each other on Between-Category trials ($F(2, 57) = 2.05$, $p > .1$), or Within-Category trials ($F(2, 57) = 2.61$, $p > .05$). The lowest mean accuracy ($M = 11.90$, $SD = 1.52$) was for HNA on Within-Category trials; this score was above chance ($t(9) = 3.94$, $p = .003$).

Continuum 4

A 2 (Group: Himba vs. English) × 2 (Trial Types: Between-Category vs. Within-Category Trials) ANOVA, with repeated measures on the second factor was carried out on the accuracy data. A significant interaction ($F(1, 57) = 17.29$, $p < .001$) was found in the context of a main effect of Trial Type ($F(1, 57) = 30.15$, $p < .001$) with superior performance on Between-Category trials but no main effect of Group ($F < 1$) (see Figure 4d). The interaction showed that the English were more accurate on Within-Category trials ($t(57) = 2.38$, $p = .02$) and the Himba on Between-Category trials ($t(57) = 2.21$, $p = .031$). However, English participants did not distinguish between the Trial Types ($t(29) = 1.22$, $p > .2$), whereas the Himba did ($t(28) = 5.71$, $p < .001$) scoring more accurately on Between-Category trials than Within-Category trials.

The lowest mean accuracy ($M = 11.38$, $SD = 3.61$) was for Himba on Within-Category trials; this score was above chance ($t(28) = 2.05$, $p = .05$).

Continuum 5

A 3 (Group: Himba Naming Agreement (HNA) vs. Himba Naming Disagreement (HND) vs. English) × 2 (Trial Types: Between-Category vs. Within-Category Trials) ANOVA, with repeated measures on the second factor, was carried out on the accuracy data. It showed a significant interaction ($F(2, 57) = 8.26$, $p < .001$) in the context of a main effect of Trial Type ($F(2, 57) = 8.41$, $p = .005$) with superior performance on Between-Category trials and of Group ($F(2, 57) = 5.96$, $p = .004$) with better performance for the Himba (see Figure 4e). Analyses of the interaction showed differences between groups only for Between-Category trials ($F(2, 59) = 15.68$, $p < .001$) and not Within-Category trials ($F < 1$). The largest difference was found between the HNA and the English Groups (Tukey HSD, $p < .001$), with a trend towards a significant difference between the HND and English Groups (Tukey HSD, $p = .056$). When looking at performance differences according to Trial Types, the only significant difference was found for HNA ($t(18) = 4.42$, $p < .001$) with an advantage for Between-Category Trials; HND ($t(10) = 1.51$, $p > .1$) and English ($t(29) = −1.53$, $p > .1$) groups performed equally well on both Trial Types. The lowest mean accuracy ($M = 10.73$, $SD = 1.64$) was for Himba on Within-Category trials; this score was above chance ($t(29) = 2.45$, $p = .021$).

Discussion

The present study examined CP with a type of stimulus not previously used to assess the effect of labels on perceptual judgments. In cross-cultural studies of colour CP, it has been found that colour categories correspond to the terms in the speaker's language (Roberson *et al.*, 2000, 2004, 2005). Indeed, Roberson *et al.* showed that colour CP for

the Himba was different from CP in another language with the same number of colour terms. For the cattle patterns, the Whorfian view predicted that it would only be the Himba who showed superiority for cross-category decisions as only they have the appropriate labels. The Whorfian view was largely upheld from the recognition accuracy for these animal stimuli.

Continua 1 and 4 gave the clearest results. For these continua, the Himba participants showed strong name agreement to those given at production of the continua. For Continua 1 and 4, the Himba showed greater recognition for cross-category trials whereas the English found the two types of trial equally easy. The pattern of results was, in Goldstone's (1994) terms one of between-category expansion in that it would seem that, to the Himba, animals from different categories look more different. One can also see, for example in Continuum 4, evidence of within-category compression in that the Himba performed worse than the English on these trials. Similar trends are in the other continua but they are difficult to be confident of in the context of the generally superior accuracy for the Himba with these animal patterns. The same pattern of cross-category advantage was seen for Continua 3 and 5 though the data were complicated by the fact that not all the Himba agreed on the names given by those who helped make those continua. However, if we considered only those Himba who showed name agreement, exactly the same pattern was found. The Himba who showed name disagreement showed category effects identical to the English participants. The clear pattern in the other four continua was less marked in Continuum 2 that only produced a trend in the same direction of the Himba showing superior performance for cross-category trials. We offer no certain explanation for the weaker effect for this continuum. However, as noted above, there are effects of contrast that could prompt the use of additional labels. We also note that the only chance score (English, Between-Category) was found on Continuum 2 and one of the within-category pairs appeared to offer easier recognition for the English participants.

It would seem that the CP found for the Himba derives from their possession of pattern categories not found in the Western participants. However, alternative explanations might be considered. Stimuli were presented to the Himba manually and it is possible that this different procedure altered overall performance. In fact, despite it being an artificial task, the Himba were generally better at the recognition of the animals than the English participants. In any case, the critical aspect of the data is not the overall level of the performance but the superiority of between-category trials, and this cannot be explained by the pictures being shown as photographs rather than on the computer. We can also rule out the possibility that the task was simply too difficult for the English observers. While there is significantly lower overall performance for the English observers for several of the continua, this is not always the case and performance levels were above chance for English participants on almost all the continua.

Another potential artifact is that of explicit naming by the Himba whereby, it could be argued, superior recognition memory could be achieved as a result of matching by name in the 2AFC. The 2AFC task is that most commonly used in cross-cultural studies of CP (Roberson *et al.*, 2000, 2004, 2005) but it is perhaps not the best. However, the same argument about labelling could have been made if the 2AFC task had been replaced by similarity judgments to determine CP (see Roberson *et al.*, 2000). We could have more clearly ruled out explicit naming by use of interference procedures (Roberson & Davidoff, 2000) but these were not possible with a Himba population. Probably a better technique would have been to employ rapid presentations that would make naming

difficult (Gilbert, Regier, Kay, & Ivry, 2006) though even then some limited naming might be possible.

In fact, there was no evidence to the experimenter of any explicit naming but it would be difficult to rule out some implicit naming from this first study of animal pattern CP. Indeed, one might not want to rule it out; activation of a perceptual difference linked to a linguistic network (i.e. implicit naming) could be critical to all CP (Roberson & Davidoff, 2000). An indication that categorical representations are activated implicitly was shown in the neuropsychological case study of Roberson *et al.* (1999). The patient demonstrated intact CP (colours and facial expressions) when using the 2AFC procedure as given to the Himba. The limited role that explicit labelling played in the task was clear because the patient was unable to name colours or facial expressions correctly, so any such attempt by him would have been detrimental to his performance.

The neuropsychological data (Roberson *et al.*, 1999) are also consonant with the view that networks activated in colour CP are not even implicitly linguistically driven but are simply perceptual (Pilling, Wiggett, Ozgen, & Davies 2003). Indeed, the introspection of observers (Kay & Kempton, 1984) is that discriminating between colours from the same or different categories is a perceptual decision. The issue concerning the underlying perceptual or linguistic mechanism for CP has, to date, been examined by use of interference techniques. In some cases, verbal interference has resulted in complete abolition of CP (Roberson *et al.*, 2000; Gilbert *et al.*, 2006) and in others less than complete abolition (Pilling *et al.*, 2003). For the case of animal pattern CP, as there is no likelihood of innately given categories, the perceptual codes underlying CP would presumably be linked to a linguistic network but that would need verification in future studies.

We regard CP for animal patterns as highly likely for the Himba. Colour categories have very limited functional significance for the Himba; this is not the case for animal patterns. However, the CP could arise from ways other than from labels. A different explanation for the origin of animal pattern CPs could be that they are derived from the function of an animal with a particular type of pattern though of course these functions are going to be highly correlated with labels. Indeed, by adulthood, it is clear from our HNA groups that the classification is intrinsically linked to the names but a distinction between the two potential origins might be possible in future testing. Even very young children look after animals in the Himba culture, so one might be able to carry out the same sort of longitudinal study as in Roberson *et al.* (2004) to investigate the development of animal pattern CP. In Roberson *et al.* (2004), the acquisition of colour CP was linked closely to the acquisition of colour terms. It might be different for animal patterns though one must stress that it would be hard to distinguish between the two explanations as pattern names might be acquired very early given their cultural significance.

In summary, the data clearly imply that labels have affected the Himba perception of animal patterns. Indeed, one could simply point to the continua in Figure 2a–e and be surprised that what appears to us to be a smooth progression could be marked with an abrupt change to any observer. The proposal that it is the acquisition of vocabulary in the language that causes the abrupt change is rather nicely confirmed by the fortuitous result from the HND groups in Continua 3 and 5 who gave no evidence for cross-category superiority. Thus, though one cannot rule out alternative explanations, the consideration of animal patterns has given further evidence in favour of linguistic relativity with respect to categorical perception.

Acknowledgements

References

Bornstein, M. H. (1975). Hue is an absolute code for young children. *Nature, 256*, 309-310.

Bornstein, M. H. (1987). Perceptual categories in vision and audition. In S. Harnad (Ed.), *Categorical perception: The groundwork of cognition* (pp. 287-300). Cambridge: Cambridge University Press.

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

Boroditsky, L., Schmidt, L. A., & Phillips, W. (2003). Sex, syntax and semantics. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind* (pp. 61-79). Cambridge, London: MIT Press.

Bowerman, M., & Choi, S. (2001). Shaping meanings for language: Universal and language-specific in the acquisition of spatial semantic categories. In M. Bowerman & S. C. Levinson (Eds.), *Language acquisition and conceptual development* (pp. 475-511). Cambridge: Cambridge University Press.

Carroll, J. B. (Ed.) (1956). *Language, thought and reality: Selected writing of Benjamin Lee Whorf*. Cambridge, MA: MIT Press.

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.

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

Eckl, A. E. (2000). Language, culture and environment. The conceptualization of Herero cattle terms. In M. Bollig & J-B. Gewald (Eds.), *People, cattle and land: Transformation of a pastoral society in southwestern Africa (History, Cultural Traditions and Innovations in Southern Africa, Vol. 13)* (pp. 401-431). Cologne: Ruediger Koeppe.

Etcoff, N. L., & Magee, J. J. (1992). Categorical perception of facial expressions. *Cognition, 44*(3), 227-240.

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

Franklin, A., & Davies, I. R. (2004). New evidence for infant colour categories. *British Journal of Developmental Psychology, 22*(3), 349-377.

Gelman, R., & Butterworth, B. (2005). Number and language: How are they related? *Trends in Cognitive Sciences, 9*, 6-10.

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

Gilbert, A. L., Regier, T., Kay, P., & Ivry, R. B. (2006). Whorf hypothesis is supported in the right visual field but not the left. *Proceeding of the National Academy of Sciences, 103*(2), 489-494.

Goldstone, R. (1994). Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology, General, 123*(2), 178-200.

Gordon, P. (2004). Numerical cognition without words: Evidence from Amazonia. *Science, 306*, 496-499.

Gumpertz, J. J. & Levinson, S. C. (Eds.). (1996). *Rethinking linguistic relativity*. Cambridge, UK: Cambridge University Press.

Harnad, S. (1987). Psychophysical and cognitive aspects of categorical perception: A critical overview. In S. Harnad (Ed.), *Categorical perception: The groundwork of cognition* (pp. 1-25). Cambridge: Cambridge University Press.

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

Kay, P., & Kempton, W. (1984). What is the Sapir-Whorf hypothesis? *American Anthropologist*, *86*(1), 65–79.

Kay, P., & Regier, T. (2003). Resolving the question of colour naming universals. *Proceedings of the National Academy of Science*, *100*(15), 9085–9089.

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

Lucy, J. A. (1992). *Language diversity and thought: A reformulation of the linguistic relativity hypothesis*. Cambridge: Cambridge University 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., Gennari, S., Shi, M., & Wang, Y. (1999). Knowing versus naming: Similarity and the linguistic categorization of artifacts. *Journal of Memory and Language*, *40*, 230–262.

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* (pp. 113–155). Cambridge, MA: MIT Press.

Namibian Government Statistics (2004). Available from: www.grnnet.gov.na/.

Nisbett, R. E., & Miyamoto, Y. (2005). The influence of culture: Holistic versus analytic perception. *Trends in Cognitive Sciences*, *9*(10), 467–473.

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*(4), 477–493.

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

Pilling, M., Wiggett, A., Ozgen, E., & Davies, I. R. L. (2003). Is color 'categorical perception' really perceptual? *Memory and Cognition*, *31*(4), 538–551.

Roberson, D., & Davidoff, J. (2000). The categorical perception of colour and facial expressions: The effect of verbal interference. *Memory and 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., Davies, I., & Shapiro, L. R. (2004). The development of color categories in two languages: A longitudinal study. *Journal of Experimental Psychology: General.*, *133*(4), 554–571.

Roberson, D., Davidoff, J., Davies, I., & Shapiro, L. R. (2005). Color categories: Evidence for the cultural relativity hypothesis. *Cognitive Psychology*, *50*, 378–441.

Roberson, D., Davidoff, J., & Shapiro, L. R. (2002). Squaring the circle: The cultural relativity of 'good' shape. *Journal of Cognition and Culture*, *2*, 29–51.

Roberson, D., Davies, I., & Davidoff, J. (2000). Color categories are not universal: Replications and new evidence from a stone-age culture. *Journal of Experimental Psychology*, *129*(3), 369–398.

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.

Turton, D. (1980). There's no such beast: Cattle and colour naming among the Mursi. *Man*, *15*(2), 320–338.