ELSEVIER

Available online at www.sciencedirect.com

SCIENCE DIRECT®

J. Experimental Child Psychology 90 (2005) 114–141

Journal of Experimental Child Psychology

www.elsevier.com/locate/jecp

Color term knowledge does not affect categorical perception of color in toddlers

Anna Franklin*, Ally Clifford, Emma Williamson, Ian Davies

Department of Psychology, University of Surrey, Guildford, Surrey, GU2 7XH, UK

Received 29 March 2004; revised 1 October 2004
Available online 19 November 2004

Abstract

Categorical perception of color is shown when colors from the same category are discriminated less easily than equivalently spaced colors that cross a category boundary. The current experiments tested various models of categorical perception. Experiment 1 tested for categorical responding in 2- to 4-year-olds, the age range for the onset establishment of color term knowledge. Experiment 2 tested for categorical responding in Himba toddlers, whose language segments the color space differently from the way in which the English language does so. Experiment 3 manipulated the conditions of the task to explore whether the categorical responding in Experiments 1 and 2 was equivalent to categorical perception. Categorical perception was shown irrespective of naming and was not stronger in those children with more developed color term knowledge. Cross-cultural differences in the extent of categorical perception were not found. These findings support universalistic models of color categorization and suggest that color term knowledge does not modify categorical perception, at least during the early stages of childhood.

Keywords: Categorical perception; Color term acquisition; Color perception; Perceptual reorganization; Linguistic relativity; Categorization

---

* Corresponding author.
E-mail address: a.franklin@surrey.ac.uk (A. Franklin).

0022-0965/$ - see front matter
doi:10.1016/j.jecp.2004.10.001

Introduction

Color is perceived categorically. The color spectrum is continuous; however, this continuum is perceived as a number of discrete categories, with the members of categories resembling each other more than they resemble members of other categories. This has been termed "categorical perception" (Harnad, 1987). Stimuli from within a category (within-category stimuli) are perceived as more similar than stimuli that straddle a category boundary (between-category stimuli), even when stimulus separation sizes for within- and between-category stimuli are equal. For example, in Fig. 1, the three stimuli A1, A2, and B are equidistant in color space, two stimuli (A1 and A2) belong to the same linguistic color category (e.g., blue), and the third stimulus (B) belongs to an adjacent linguistic category (e.g., green).

Categorical perception is shown when the stimulus pair A2–B1 is discriminated faster, more easily, or more accurately than the stimulus pair A1–A2. Categorical perception of color has been evidenced using same–different judgment, recognition memory, and two-alternative forced-choice tasks (2-AFCs) (e.g., Bornstein & Korda, 1984; Pilling, Wiggett, Özgen, & Davies, 2003; Roberson & Davidoff, 2000; Uchikawa & Shinoda, 1996).

The origin of categorical perception of color is under debate. On the one hand, universalists argue that categorical perception is "hardwired" into the visual system and that categorical perception of color is an innate, universal, and perceptual effect. On the other hand, linguistic relativists argue that categorical perception of color is constructed through language—an idea that has its roots in the Sapir–Whorf hypothesis that language determines thought (Sapir, 1921; Whorf, 1956). Language "warps" perceptual space, creating compression of within-category perceptual space and expansion of between-category perceptual space.

The universalistic theory leads to the prediction that categorical perception should be shown during infancy—before color terms are learned. Studies that show categorical perception of color in 4-month-olds (e.g., Bornstein, Kessen, & Weiskopf, 1976; Franklin & Davies, 2004) support this prediction. Using a habituation technique and monochromatic lights, Bornstein and colleagues (1976) showed that infants respond

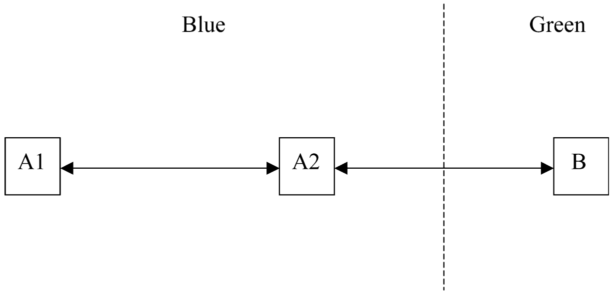

Fig. 1. Diagram showing the classic categorical perception design. The boxes represent stimuli, the arrows represent stimulus separations, and the dashed line represents a category boundary.

categorically across blue–green, yellow–green, and yellow–red boundaries. Using a novelty preference technique and the Munsell color metric,[1] Franklin and Davies (2004) showed that infants respond categorically across blue–green and blue–purple hue boundaries and across a pink–red boundary defined by differences in lightness and saturation.

The linguistic relativity theory leads to the prediction that categorical perception will vary as language varies. Cross-cultural studies of categorical perception in adults (e.g., Kay & Kempton, 1984; Pilling & Davies, in press; Roberson, Davies, & Davidoff, 2000) support this prediction. For example, Roberson and colleagues (2000) tested for categorical perception (using a 2-AFC task) in two populations whose languages segment the color space differently from each other. Whereas English distinguishes between *blue* and *green*, the Berinmo language (Papua New Guinea) does not. In contrast, Berinmo makes a distinction (*nol/wor*) that English does not. Category effects were shown by English and Berinmo speakers, but only at the *blue–green* boundary for the English speakers and only at the *nol–wor* boundary for the Berinmo speakers.

In summary, there is evidence to support both sides of the debate. On the one hand, prelinguistic infants show categorical perception across a range of color boundaries. On the other hand, cross-cultural differences in categorical perception are found. These two streams of evidence seem to contradict each other. How is it possible that prelinguistic infants show categorical perception, whereas adults do not show categorical perception when their language does not mark the category boundary? The apparent contradiction is based on two crucial, but untested, assumptions: (a) that infant categorical perception is hardwired and universal and (b) that infant categorical perception is equivalent to adult categorical perception. Infant color categorical perception has been found in British and American infants, but no such tests have been made on infants from other language groups. The behavioral markers of infant categorical perception are based on either dishabituation (Bornstein et al., 1976) or novelty preference (Franklin & Davies, 2004), whereas adult categorical perception is evidenced typically using 2-AFC discrimination tasks (e.g., Roberson et al., 2000). Thus, infant and adult behaviors may be based on different processes. However, if these two crucial assumptions are accepted, it seems to imply that categorical perception disappears unless it is supported by linguistic distinctions; some kind of "perceptual reorganization" must occur.

Infant categorical perception: Universal and hardwired?

The cross-cultural variation in categorical perception in adults is not necessarily due to language. The cross-cultural studies show an association between language and categorical perception, but they do not show that language causes the variation. The cross-cultural differences in categorical perception could predate language

---

[1] The Munsell metric is a perceptually uniform space that produces highly reliable standardized colors. The Munsell color space has three dimensions: Hue, Value (lightness), and Chroma (colorfulness, like saturation).

acquisition. For example, perhaps Berinmo infants would not show categorical perception across the *blue–green* boundary but would show categorical perception across the *nol–wor* boundary. Moreover, it could be these prelinguistic differences that lead to different color lexicons. However, although this argument is valid logically, its plausibility is low because it questions a major tenet of vision science—the universality of early visual processes—without supporting evidence.

The first stages of chromatic processing are based on three different classes of cones with different spectral sensitivities. Processing within the retina recombines the cone signals into three opponent channels that are carried to the visual cortex on anatomically separate pathways (e.g., Gegenfurtner & Kiper, 2003). The opponent signals are based on the receptive field structure of ganglion and thalamic cells. This basic structure is assumed to be common to all individuals with normal color vision. However, this structure is not fully developed at birth. Rather, retinal and receptive field organization develops, probably over several years (e.g., Atkinson, 1984). Neuronal growth that leads to receptive field structure could be completely hardwired, there could be a random component to it, or there could be a degree of environmental tuning to it (MacLeod, 2003; Yendrikhovskij, 2001). The latter possibility implies that different chromatic environments could lead to different chromatic processing structures. In extremis, restricted chromatic environments during the first year lead to defective color vision and abnormal color constancy in macaque monkeys (Sugita, 2004). However, even if chromatic tuning occurs, it is not clear what chromatic regularities neuronal growth would respond to, and these regularities could be universal (Shepard, 1992).

There are variations in the optical properties of the eye and retinal sensitivity that could produce differences in color vision. Preretinal filtering varies due to different optical densities of the lens and macula pigmentation (Bornstein, 1973), and there are variations in the spectral sensitivities of cone photopigments (e.g., Jameson, Highnote, & Wasserman, 2001). However, there is no evidence that these variations covary with language, culture, or region. Webster and colleagues (2002) did report systematic differences in the identification of focal and unique hues between Indian and American populations, and physiological, environmental, and cultural reasons for the differences were considered. However, the origin of this difference is not clear.

The second part of the assumption is that infant categorical perception is hardwired. The brain structures responsible for color categories are not known, but they almost certainly involve cortical structures beyond the visual cortex. Perhaps these structures are tuned by the chromatic environment such that by 4 months of age, adultlike categorical perception is shown. As with our speculations about tuning of earlier chromatic stages, there is no direct evidence that this occurs, nor is it clear what chromatic regularities would influence tuning. However, one possibility is that infant environments in the industrialized world are dominated by artifacts (e.g., toys, clothes, pictures) in saturated primary colors that are close to category prototypes (Rosch, 1972). Caregivers may also emphasize these prototypes by, for instance, bringing them to the attention of infants. Perhaps such processes tune category formation, leading to categorical perception by 4 months of age. In contrast, infants in tribal societies, such as the Berinmo, have chromatic environments based mostly on natural objects that tend to be in more muted colors. This could lead either to less salient categories or even to different categories.

Equivalence of infant and adult categorical perception

Both infants and adults behave as though there is something pertinent about categorical differences of colors. However, even on tasks that seem to be formally equivalent, such as adult 2-AFC and infant novelty preference, there are differences that call their equivalence into question. The formal equivalence is that both infant and adult tasks involve 2-AFCs. The infants are familiarized to a standard color and then confronted with the standard paired with a novel color, either from the same category as the standard or from an adjacent category. Categorical perception is evidenced by infants looking more at the categorically different test color than at the same category test color. In adult 2-AFC, a target color is presented, followed after an interval by two test colors: One identical to the target and one different (the foil). As with infants, on some trials the different color is from the same category as the target, and on some trials the different color is from an adjacent category. For adults, accuracy is the performance measure; for infants, the degree of novelty preference, as indicated by direction of gaze, is the index of performance. Even assuming that the two measures are equivalent, adult and infant patterns of behavior indicating categorical perception differ. Adults are more accurate for between-category pairs than for within-category pairs, but their within-category accuracy is well above chance (e.g., Pilling et al., 2003). In contrast, infants show no novelty preference for within-category pairs; they seem to require a categorical change to engage their interest. Although this is sufficient to suggest that infants have something like color categories, it also suggests that they might not yet have developed their adult form. Beyond their formal equivalence, the tasks differ between infants and adults. First, adults are explicitly told that they have to choose the color they think is identical to the target, whereas infants are (necessarily) not instructed. Second, adults are likely to bring to bear on the task a repertoire of strategies that are not available to infants. Perhaps most crucially, adults may use verbal labels for the stimuli to help them remember the target color across the interval (we return to this in the next section).

These differences between adult and infant measures of categorical perception suggest that it would be useful to test for categorical perception prelinguistically using tasks as similar to the adult tests as possible. Studies that test for categorical perception as soon as children are able to be instructed, using tasks such as 2-AFC discrimination, would bridge the gap between the current infant and adult studies of categorical perception.

The perceptual basis for color categorical perception

So far, we have followed conventional use by using the term "categorical perception." However, although it is clear that there are categorical influences on responding, the perceptual basis for these effects is inferred only indirectly. Most tasks used to measure categorical perception require the use of memory, and for adults performance might also be influenced by labeling. The evidence is consistent with categorical perception arising from perceptual processes, but the evidence is also consistent with it being due to labeling or arising from memory processes.

Labeling could produce apparent categorical perception as follows. If the target color in a 2-AFC task is labeled and the label is retained across the interval to be compared with the labels given to the two test stimuli (target plus foil), then this would support accurate choices when the foil and target are in different categories (e.g., blue–green) but not when they are in the same category (e.g., blue–blue). A simple model based on naming reliability predicts between-category performance well but underestimates within-category performance for delayed 2-AFC and same–different tasks (Pilling et al., 2003). Consistent with this labeling account, Roberson and Davidoff (2000) found that categorical perception is eliminated by verbal interference (presentation of a list of words that had to be read), but not visual interference (presentation of a stationary curved line that had to be tracked visually), during the interstimulus interval (ISI) of a successive 2-AFC task. Categorical perception in infants cannot be due to verbal labeling, but verbal labeling could explain the cross-cultural differences in categorical perception seen in adults. For example, perhaps the Berinmo did not show categorical perception for blue–green because they both have the same label (Munnich & Landau, 2003). With the use of nonlinguistic tasks, cross-cultural differences might not be seen and the underlying perceptual categorization of color might be revealed to be universal.

Categorical perception-like effects could also arise from "distortions" in memory traces. For example, Huttenlocher, Hedges, and Vevea (2000) argued that as uncertainty about a target stimulus increases with factors such as noisy stimuli, reduced attention, and increasing retention periods, it pays to bias judgments toward the category prototype. Because within-category stimuli have the same prototype and between-category stimuli have different prototypes, a shift toward the prototype would make the representations of target and foil more different for between-category pairs but more similar for within-category pairs, mimicking categorical perception. There is evidence that memory errors can be biased systematically (Davies, Özgen, Pilling, & Riddett, 2004; Petzold & Sharpe, 1998; Prinzmetal, Amiri, Edwards, & Allen, 1998). However, Pilling and colleagues (2003) found that color categorical perception and memory bias occurred but were dissociated. Moreover, color categorical perception is found using simultaneous same–different tasks (Özgen & Davies, 2003) and visual search tasks (e.g., Daoutis, Franklin, Riddett, Clifford, & Davies, 2004; Daoutis, Pilling, & Davies, 2004; Franklin, Pilling, & Davies, 2004) where the memory load is minimal and labeling is of no use.

It is quite possible that "categorical responding" could arise from perception, and from memory and labeling, and that these different sources are more or less active depending on the particular task and the strategies deployed. If the concern is whether genuine categorical perception is occurring, then studies of categorical perception need to include tasks that can either isolate perceptual processes or exclude memory and labeling mechanisms.

Perceptual reorganization

A third theory that combines the ideas of universalism and linguistic relativism could resolve the conflict. The resulting theory, the theory of "perceptual reorganization," postulates that there is an innate predisposition for category boundaries at certain points in the color space but that language learning modifies the location and extent of categorical perception, reorganizing the representation of perceptual color space. Here the process of learning color terms would draw attention to the similarities of colors given the same term and would draw attention to the differences of colors given different terms. If color perception is "plastic," then this process could cause compression of perceptual space for the areas of color space that are given the same term and could cause expansion of perceptual space for the areas of color space where there is a linguistic boundary. Such perceptual learning could create categorical perception across newly learned boundaries and could attenuate or eliminate it from within-category regions, altering the structure of perceptual categorization. Thus, categorical perception shown during infancy may later be lost if a language does not mark the boundary, and categorical perception may be accentuated and sharpened if a language does mark the boundary. Finally, categorical perception not present during infancy may later be acquired if a language does mark the distinction between two color regions.

There is some evidence that supports the model. The flexibility of categorical perception in general is supported by category training studies (Goldstone, 1994; Goldstone, Lippa, & Shiffrin, 2001; Guenther, Husain, Cohen, & Shinn-Cunningham, 1999; Livingston, Andrews, & Harnad, 1998; Özgen & Davies, 2002). For example, Özgen and Davies (2002) provided clear evidence of the plasticity of color discrimination and the modifiability of category structure. First, they showed that color discrimination could be improved by training and that the learning was restricted to the training stimuli. Second, they showed that learning to subdivide a preexisting basic color category induced categorical perception across the new boundary. The boundary was at the center of either the blue or green category, and before training discrimination in this region was the minimum for the category. Training reversed this pattern, with the old category center now having peak discriminability.

The idea that universal hardwiring is at some point molded by linguistic or environmental input can be found in other domains. For example, there is evidence that some speech contrasts discriminated by infants are later lost if a language does not encode them (Werker & Lalonde, 1988; Werker & Tees, 1983, 1984). Hespos and Spelke (2004) showed that 5-month-olds reared in an English-speaking environment make a conceptual distinction that is marked by the Korean language, whereas English-speaking adults do not make this distinction. The mechanisms that are at play in these studies are not necessarily the same (Bloom, 2004) and are not necessarily the mechanisms that would be needed for perceptual reorganization of color. However, these studies, as well as Goldstone's (1994) and Özgen and Davies's (2002) studies, may suggest that there is a degree of plasticity to discrimination and categorization (see also Fahle & Poggio, 2002).

To summarize, there is no current consensus about the origin or nature of categorical perception. Prelinguistic infants show categorical perception, but it is not clear whether the category effect is equivalent to categorical perception in adult studies or whether the category effect is perceptual. It is also not clear whether prelinguistic categorical perception is universal. Cross-cultural differences in categorical perception exist, but it is not clear whether cross-cultural differences are a result of language learning or whether these differences exist before language. It is also not clear whether these differences are truly perceptual or whether these differences are due to verbal labeling. Therefore, there are various issues that need to be resolved. First, can categorical perception be shown prelinguistically using tasks equivalent to the tasks in the adult studies, and what is the impact of language learning on the extent of categorical perception? Second, if categorical perception is shown prelinguistically, is it universal? Third, is categorical perception truly perceptual? Before these issues are resolved, a clear understanding of the origin and nature of categorical perception cannot be reached.

Overview of the current experiments

The set of experiments presented here combine developmental and cross-cultural approaches to investigate each of these issues further. In Experiment 1, category effects are tested in 2- and 3-year-olds at different stages of color term knowledge. The three boundaries of Franklin and Davies's (2004) investigation of categorical perception of color in infants are tested. A 2-AFC task (equivalent to the task used in adult studies of categorical perception) is used. Experiment 2 uses the same technique as Experiment 1 and tests for category effects in a group of toddlers from rural Namibia who have not yet learned color terms and whose parental language has 5 basic color terms (as opposed to the 11 terms in English).

Each of the three models of categorical perception would predict different results for Experiments 1 and 2. The universalistic model would predict that all toddlers, irrespective of color term knowledge or parental language, would respond categorically. The linguistic relativity model would predict that only the English toddlers in Experiment 1, who have learned to linguistically mark category boundaries, would respond categorically. The perceptual reorganization model would predict that all toddlers, irrespective of color term knowledge or parental language, would respond categorically but that the English toddlers in Experiment 1 (who have learned their color terms) would show a stronger category effect.

Experiment 3 looks at the nature of the category effect by varying the memory load of the task to ascertain whether the category effect found is based on a memory process. If the locus of the category effect is memory, then the size of the category effect should decrease as the memory component of the task is reduced. If the locus of the category effect is perception, then a category effect should be found even when there is no memory component to the task. Until it is clear whether the locus of any category effect found is perceptual, the term "categorical responding" will be used instead of "categorical perception."

Experiment 1: The effect of color term acquisition on categorical responding to color in toddlers

Early studies of color term knowledge in children reported that color terms were not reliably acquired until around 4–7 years of age (Bornstein, 1985). However, more recent studies have shown that color terms can be reliably acquired at as young as 2 years of age (e.g., Andrick & Tager-Flusberg, 1986; Shatz, Behrend, Gelman, & Ebeling, 1996). Reliable knowledge of the first nine basic color terms is established at around 3 years of age, with the acquisition of brown and gray being established roughly 6 to 9 months later (Pitchford & Mullen, 2002). The main aim of the Experiment 1 was to test for category effects during this stage in development.

We test for category effects using a 2-AFC task, that is, a task that is used to test for categorical perception in adults. However, the adult version of this task would not be suitable for young children; toddlers would find it difficult to understand the task and would find it hard to sustain motivation and concentration over several trials. Therefore, the adult 2-AFC task was modified for use with young children. Accuracy for between- and within-category target–foil pairs was assessed. Higher accuracy for between-category pairs was taken as an index of categorical responding. The size of the category effect for each boundary (blue–green, blue–purple, or pink–red) was assessed and compared. The impact of linguistic categorization on the extent of the category effect was investigated. Each child named the test stimuli used in the main experiment. In addition, each child completed a naming and comprehension task using 11 basic focal stimuli.

The naming task that uses the test stimuli was intended to reflect the clarity of the linguistic boundary. Individual naming patterns of the test stimuli were classified as having no linguistic boundary, a correct linguistic boundary, or a reversed linguistic boundary. If linguistic categorization creates the category effect (linguistic relativity model), then (a) those children with no linguistic boundary should show no category effect, (b) those children with a correct linguistic boundary should show between-category facilitation, and (c) those children with a reversed linguistic boundary should show within-category facilitation. If linguistic categorization amplifies the category effect (perceptual reorganization model), then (a) those children with no linguistic boundary should show a weaker category effect than those children with a boundary, (b) those children with a correct linguistic boundary should show between-category facilitation; and (c) those children with a reversed linguistic boundary should show within-category facilitation. If linguistic categorization has no impact on the category effect (universalistic model), then all children would respond categorically to the same extent, irrespective of their pattern of naming.

The naming and comprehension tasks that use focal stimuli were intended to reflect more general color term knowledge and color fluency. If categorical perception and language are linked, we may expect the size of the category effect to be related to the number of focal stimuli correctly named and identified.

To summarize, there are three main questions. First, do children of this age respond categorically across the same boundaries as the infants? Second, is it only the children who linguistically mark the boundaries that show categorical responding, or is categorical responding greater if children linguistically mark the boundaries? Alternatively, is categorical responding shown irrespective of color term knowledge?

Method

Participants

A total of 60 English-speaking children (33 boys and 27 girls) between 26 and 47 months of age (mean age = 36 months, *SD* = 6.1) were recruited from local nurseries. Children were not tested for color-blindness because children of this age find such tests to be difficult to understand and complete.

Stimuli and design

Two identical bear figures (Bear A and Bear B) were cut out of white cardboard, with the bear outline and facial features drawn with black pen. Colored stimuli were cut out in the shape of a sweater that could be placed on the bears. All sweaters were the same shape and were mounted on cardboard. Two types of sweaters were made.

Focal stimuli.

Sweaters were made from Color-Aid paper that provided good examples of the 11 focal colors (with Color-Aid codes in parentheses): black (BLACK), white (WHITE), red (RO Hue), green (G Hue), yellow (Y Hue), blue (B Hue), brown (O S3), pink (R T4), purple (V Hue), orange (YO Hue), and gray (GRAY 4). Two sweaters of each color were made.

Test stimuli.

The test stimuli sweaters were made from Munsell Glossy paper, and two instances of each color were used. There were three sets of test stimuli—blue–green, blue–purple, and pink–red—with three stimuli per set. Test stimuli for the blue–green and blue–purple sets varied only in Munsell hue; Munsell Chroma (colorfulness, like saturation) and Munsell Value (lightness) were kept constant. Test stimuli for the pink–red set varied in Munsell Chroma and Munsell Value; Munsell Hue was kept constant. For each set, there was a within-category pair and a between-category pair, and the separations of within- and between-category pairs were equated in Munsell units and were also equal in color space units (ΔE) of another perceptually uniform color space (the CIE ($L^*u^*v^*$) color space). The location of the category boundaries was confirmed by adult naming and similarity judgments (Franklin & Davies, 2004). The Munsell codes and categorical status of stimulus pairs are shown in Fig. 2.[2]

---

[2] Blue–green and blue–purple separations were around 30 units in CIE color space (30 ΔE), and pink–red separations were around 20 units (20 ΔE) (for details of the CIE color space, see Davies & Franklin, 2002). Therefore, if accuracy is based on separation sizes, children should be more accurate for the blue–green and blue–purple sets.

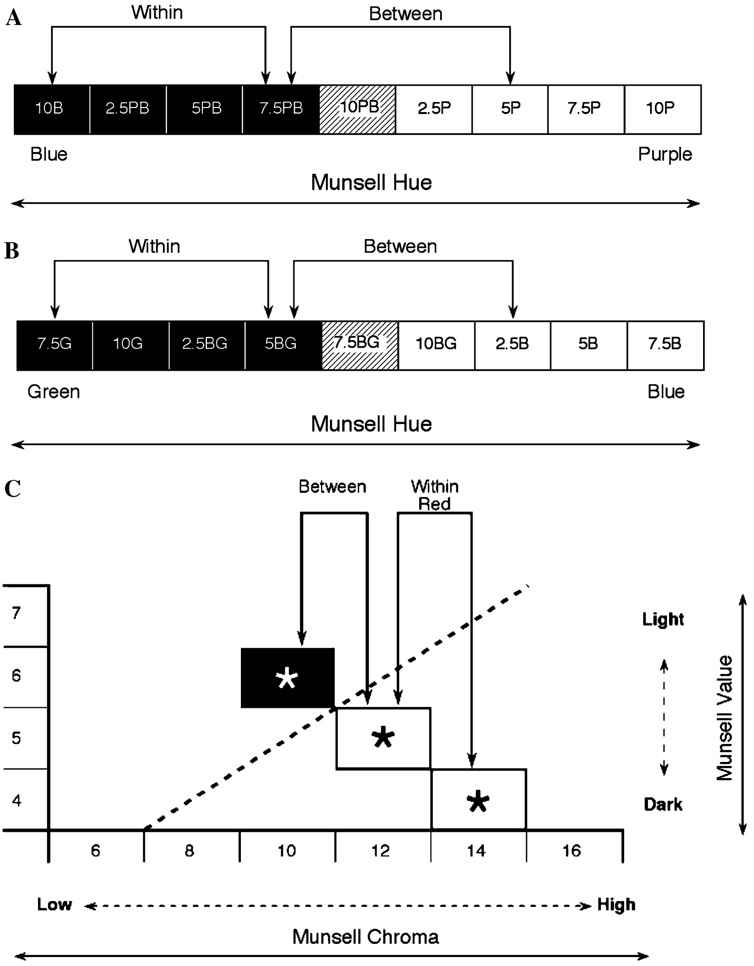

Fig. 2. Munsell codes, categorical status, and Munsell distances of the stimuli of the experimental pairs used in Experiment 1. The categorical relationships (within or between) of the experimental pairs are shown for blue–purple (A), blue–green (B), and hue boundaries and a pink–red lightness saturation boundary (C). In panel A, Chroma = 3 and Value = 10. In panel B, Chroma = 5 and Value = 10. In panel C, Hue = 5R.

The procedure was conducted under standardized lighting conditions that simulate natural daylight (D65, 6500 K, at 810–1880 lux) with the use of a GretagMacbeth lamp. These lighting conditions are necessary for the uniformity of the Munsell system to be maintained (Davies & Franklin, 2002).

Procedure

All children completed the training, comprehension, and naming task using the focal sweaters. For the 2-AFC task, children were tested on only one of the three test sets (blue–green, blue–purple, or pink–red). Then 20 children were randomly allocated to each of these sets, and the three groups were matched in age (blue–green set mean age = 36 months, *SD* = 5.1; blue–purple set mean age = 36 months, *SD* = 6.3; pink–red set mean age = 36 months, *SD* = 6.9). All children completed the 2-AFC task using the within and between pairs of their allocated sets.

Training

The aim of the training session was to show each child that when Bear A wore a particular sweater, Bear B also wore the identically colored sweater.[3] The two card bears were placed flat on the table in front of the child, who was encouraged to give each bear a name. The child was told that Bear A has lots of colored sweaters, and the 11 focal sweaters were randomly laid out above Bear A. The child was then told that Bear B also has lots of colored sweaters, and the other set of 11 focal sweaters were randomly laid out above Bear B. A sweater from Bear A's set was randomly chosen and placed on Bear A. The child was told that if Bear A wears this sweater, then Bear B also wears this sweater, and so the corresponding sweater was picked out of Bear B's set and placed on Bear B. This was repeated three times with different colored sweaters. Another sweater was chosen and placed on Bear A, but this time Bear B's set of sweaters was covered with white cardboard. The child was allowed to inspect the sweater, and after 5 s of stimulus presentation, Bear A and its sweater were covered. After a further 5-s delay, Bear B's set of sweaters were uncovered and the child was asked to find the same sweater for Bear B out of the set of 11 sweaters. After the child had made a choice, Bear A was uncovered and the child was encouraged to evaluate his or her response. If the choice was correct, then the child was praised; if the choice was incorrect, then the child was encouraged to amend his or her choice. This was repeated until a correct response was given three times and the experimenter was sure that the task was understood by the child. All of the children passed this training.

Two-alternative forced-choice task

Test stimuli sweaters were used in a 2-AFC task. The procedure and goal of finding the matching sweater for the other bear was the same as in the training task except that the child was given a choice of only two sweaters: an incorrect choice (foil) and a correct choice (target) identical to Bear A's sweater. The categorical relationship (between category/within category) of the incorrect and correct choices was manipulated (for stimulus pairs, see Fig. 3). The procedure was conducted for both within- and between-category stimulus pairs four times each. For two of these four trials, one stimulus was the target and one stimulus was the foil; for the remaining two trials, this target/foil allocation was reversed. Therefore, the child made a total of eight judgments, and the order was randomized for each child. The child was then

---

[3] Glucksberg, Hay, and Danks (1976) showed that $2\frac{1}{2}$-year-olds can understand the terms "same" and "different." The term "different" was not used in the procedure at all, and the training phase ensured that the child understood the meaning of the word "same."

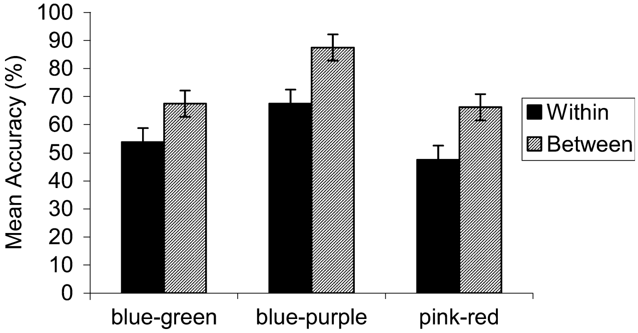

Fig. 3. Mean accuracy (±1 *SE*) for within- and between-category pairs for the blue–purple, blue–green, and pink–red sets.

presented with each of the stimuli individually and was asked to name the color of the sweater.

Comprehension and naming of focals

The focal sweaters were laid out above Bear A, and the child was asked to put a specified colored sweater on Bear A. The color of the chosen sweater was recorded, and the sweater was taken off the bear and replaced above. This was repeated until the child had been asked to place all of the colored sweaters on the bear. Each of the focal sweaters was then presented individually and in random order, and the child was asked to name the color of the sweater. The 2-AFC task was always presented first, and the naming task and comprehension task were presented in a randomized order.

Results

Category effect for the three boundaries?

For each child, the number of correct identifications on the 2-AFC task was calculated when the choice was between two stimuli of the same category (within-category pairs) and when the choice was between stimuli of different categories (between-category pairs). The maximum number of correct identifications was four for within-category pairs and four for between-category pairs. Fig. 3 gives the accuracy scores (percentage correct) for within and between pairs for the blue–purple, blue–green, and pink–red set. It appears that between-category accuracy was higher than within-category accuracy for all conditions.

This was supported by an analysis of variance (ANOVA) looking at the effects of category (between or within) and set (blue–purple, blue–green, or pink–red). Category and set were repeated measures. Accuracy was greater for between-category pairs (mean = 73.75%, $SD = 22.75$) than for within-category pairs (mean = 56.25%, $SD = 30.75$), $F(1, 57) = 18.64$, $MSE = 0.79$, $p < .001$. Overall accuracy varied across set type (blue–purple set mean = 77.50%, $SD = 21.31$; blue–green set mean = 60.62%,

$SD = 19.14$; pink–red set mean = 56.87%, $SD = 21.26$), $F(2, 57) = 5.69$, $MSE = 1.36$, $p < .01$. Accuracy was above chance (50%) for the blue–green set, $t(19) = 2.48$, $p < .005$, and the blue–purple set, $t(19) = 5.77$, $p < .001$, but not the pink–red set, $t(19) = 1.45, p = .16$. Bonferroni post hoc tests revealed significant differences in accuracy[4] between the blue–green set and the blue–purple set, $t(18) = 2.59$, $p < .05$, and between the blue–purple set and the pink–red set, $t(18) = 3.16$, $p < .005$, but not between the blue–green set and the pink–red set, $t = 0.58$, $p = .57$. The category by set interaction was not significant, $F(2, 57) = 0.22$, $p = .80$. Therefore, categorical perception was demonstrated for all sets.

Effect of naming on the size of the category effect

Table 1 gives the naming frequencies for each of the 2-AFC stimuli for all three sets. Overall, the majority name is 70% to 80% in agreement with the intended linguistic boundary for the blue–green and blue–purple sets. For the pink–red set, overall agreement was lower. Agreement on a term for Stimulus A1 was high, and there was 60% to 70% agreement on the term for Stimulus B. However, agreement for Stimulus A2 was low, with only 45% calling the stimulus red.

The accuracy scores for all sets were combined for an investigation of the effects of language because the sample sizes were not sufficient to look at the effect of language for each boundary separately. Individuals' naming patterns were analyzed to investigate the effect of linguistic categorization on the category effect. If individuals gave A1 and A2 the same term and gave B a different term, then they were classified as having a between-category linguistic boundary (boundary). An "I don't know" response was also counted as a term.[5] Therefore, two patterns of naming qualified: (A1: name 1; A2: name 1; B: name 2) and (A1: name 1; A2: name 1; B: "I don't know"). If individuals gave A2 and B the same term and gave A1 a different term, then they were classified as having a within-category linguistic boundary (reverse boundary). Again, an "I don't know" response was counted as a term. Therefore, two patterns of naming qualified: (A1: name 2; A2: name 1; B: name 1) and (A1: "I don't know"; A2: name 1; B: name 1). If individuals gave A1, A2, and B the same term, they were classified as having no linguistic boundary. Again, "I don't know" was counted as a term. Therefore, two patterns of naming qualified: (A1: name 1; A2: name 1; B: name 1) and (A1: "I don't know"; A2: "I don't know"; B: "I don't know").

In total, 37 children had a linguistic boundary, 14 had a reverse linguistic boundary, and 9 had no linguistic boundary. Fig. 4 shows mean within- and between-cat-

---

[4] Accuracy was higher for the blue–purple set (separations around 30 ΔE) than for either the blue–green set (30 ΔE) or the pink–red set (20 ΔE). Accuracy for the blue–green set (30 ΔE) and pink–red set (20 ΔE) did not differ. Therefore, accuracy was not based entirely on separation sizes.

[5] An "I don't know" response was counted as a term because it indicates that the stimulus is not regarded as belonging to a linguistic category and, therefore, has implications for the location of the linguistic category boundary.

Table 1
Frequencies (percentages) of the color terms offered for the stimuli of each set

Set	Stimulus type: Munsell code	Color term	Percentage of children offering the term
Blue–purple			
	A1: 10B 3/10	**Blue**	85
		I don't know	15
	A2: 7.5PB 3/10	**Blue**	70
		Purple	15
		I don't know	15
	B: 5P 3/10	**Purple**	85
		I don't know	10
		Green	5
Green–blue			
	A1: 7.5G 5/10	**Green**	80
		I don't know	20
	A2: 5BG 5/10	**Green**	70
		Blue	10
		Purple	5
		Red	5
		I don't know	10
	B: 2.5B 5/10	**Blue**	80
		Purple	5
		Red	5
		I don't know	10
Red–pink			
	A1: 5R 4/14	**Red**	85
		Pink	5
		I don't know	10
	A2: 5R 5/12	Red	45
		Pink	35
		I don't know	20
	B: 5R 6/10	**Pink**	65
		I don't know	20
		Red	10
		Orange	5

Note. The most common response is indicated in bold.

egory accuracy (combined across sets) for the three linguistic types. It appears that the size of the category effect does not differ for the three linguistic types.

This is supported by an ANOVA with linguistic categorization (boundary, reverse boundary, or no boundary) and category (within or between) as factors. Category was a repeated measures factor. Accuracy did not vary with linguistic categorization (mean accuracy for boundary group = 67.9%, $SD = 21.6$; mean accuracy for reverse boundary group = 64.3, $SD = 20.1$; mean accuracy for no boundary group = 54.2%, $SD = 26.5$), $F(2, 57) = 1.42$, $MSE = 969.27$, $p = 0.25$. The category effect, $F(1, 57) = 18.64$, $MSE = 479.68$, $p < .001$, also did not vary with linguistic categorization, $F(2, 57) = 1.01$, $MSE = 479.68$, $p = .37$.

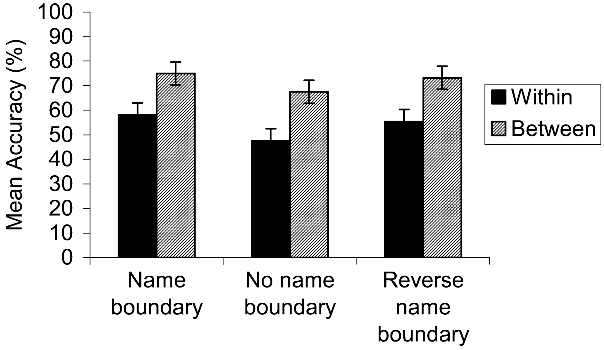

Fig. 4. Mean accuracy (±1 *SE*) for within- and between-category pairs for the three types of linguistic categorization.

Color term acquisition

Color term acquisition was analyzed at four stages: between 2 and 2.5 years of age, between 2.5 and 3 years of age, between 3 and 3.5 years of age, and between 4 and 4.5 years of age. In total, 11 children were between 2 and 2.5 years of age, 19 were between 2.5 and 3 years of age, 17 were between 3 and 3.5 years of age, and 13 were between 3.5 and 4 years of age. Table 2 and 3 give the percentages of correct responses for each focal color.

Fig. 5 shows the mean numbers of correct names given on the naming task and the mean numbers of correctly identified stimuli on the comprehension task for the four age bands. It can be seen that children's naming and comprehension of the 11 focal colors improved from a mean of approximately 6 terms at 2–2.5 years of age to a mean of approximately 10.5 terms at 3 years of age, and there was no difference between naming and comprehension. That there was no difference between naming and comprehension is unusual given that, in studies using similar tasks,

Table 2
Percentages correctly named on naming task for each age band and for all ages

	2–2.5 years (n = 11)	2.5–3 years (n = 19)	3–3.5 years (n = 17)	3.5–4 years (n = 13)	Overall (N = 60)
Black	63.6	78.9	94.1	92.3	83.3
White	72.7	68.4	82.4	100.0	80.0
Red	63.6	73.7	100.0	100.0	85.0
Green	63.6	73.7	100.0	100.0	85.0
Yellow	54.5	84.2	100.0	100.0	86.7
Blue	54.5	84.2	100.0	100.0	86.7
Orange	36.4	68.4	100.0	92.3	76.7
Pink	45.5	78.9	100.0	100.0	83.3
Purple	36.4	68.4	100.0	100.0	78.3
Brown	45.5	47.4	94.1	84.6	68.3
Gray	27.3	26.3	76.5	76.9	51.7

Table 3
Percentages correctly identified on comprehension task for each age band and for all ages

	2–2.5 years ($n = 11$)	2.5–3 years ($n = 19$)	3–3.5 years ($n = 17$)	3.5–4 years ($n = 13$)	Overall ($N = 60$)
Black	72.7	78.9	88.2	92.3	83.3
White	72.7	84.2	88.2	100.0	86.7
Red	63.6	78.9	94.1	100.0	85.0
Green	54.5	68.4	94.1	100.0	80.0
Yellow	54.5	78.9	100.0	100.0	85.0
Blue	63.6	89.5	100.0	100.0	90.0
Orange	45.5	73.7	94.1	100.0	80.0
Pink	45.5	78.9	100.0	100.0	83.3
Purple	54.5	68.4	100.0	100.0	81.7
Brown	36.4	47.4	82.4	69.2	60.0
Gray	45.5	52.6	76.5	92.3	66.7

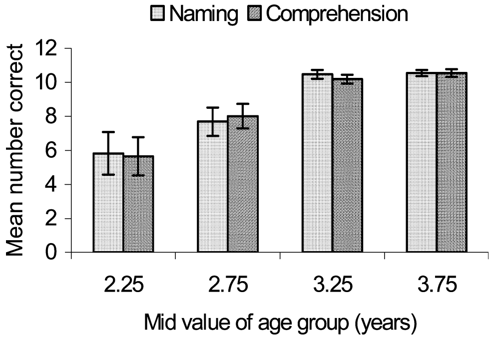

Fig. 5. Mean numbers correct on naming and comprehension tasks for each age band. Bars represent ±1 *SE*.

accuracy is usually higher for naming than for comprehension. It is unclear why this was not the case in the current experiment.

The relation between general color term fluency and the extent of the category effect was explored. This was the second test of the relation between color term acquisition and categorization. A general color term fluency index was calculated by averaging the mean number of colors named and the mean number of colors identified. A categorical effect index was calculated by subtracting the within-category score from the between-category score (on the 2-AFC task) for each child. A score higher than 0 indicates a categorical effect; that is, between-category accuracy is greater than within-category accuracy. A score of 0 indicates no categorical effect; that is, within-category accuracy and between-category accuracy are equal. A score lower than 0 indicates a reversed categorical effect; that is, within-category accuracy is greater than between-category accuracy. The relation between the general color term fluency and categorical effect indexes was significant but negative, although the r value was only -0.30 ($p < .05$).

Discussion

Toddlers responded categorically across all three boundaries, and the strength of this category effect was the same for all three boundaries. The extent of the category effect was not affected by linguistic categorization. Even though this conclusion relies on a null result, it is clear from the pattern of means in Fig. 4 that the results of the current experiment could not support the linguistic relativity model. The relation between the size of the category effect and general color term knowledge was significant but negative; increased color term knowledge actually weakened the category effect. Again, these results could not support the linguistic relativity model. The results of Experiment 1 offer support for the universalistic argument that categorical perception can be found prelinguistically. Because the task used to test for categorical perception was the same as in adult studies of categorical perception, the results also support the argument that prelinguistic categorical perception is equivalent to categorical perception in adults. The implications of the results are discussed further in the General discussion later.

Experiment 2: Categorical responding in English and Himba toddlers

Experiment 1 investigated categorical responding in English toddlers. The toddlers showed a category effect for blue–green, blue–purple, and pink–red boundaries. Experiment 2 investigated categorical perception effects in toddlers in a different population-the Himba. The Himba live in the Kaokoveld region of northwestern Namibia, which is remote, mountainous, sparsely vegetated, and arid. The Himba lead a traditional, pastoral, and nomadic life and make up less than 1% of the Namibian population (Malan, 1995). The language has five basic color terms: *oshiserandu* (denotes English red and some pinks and oranges), *oshidumbu* (denotes English beige and yellow), *oshizoozu* (denotes English black and other darker colors), *oshivapa* (denotes English white and some lighter colors), and *oshiburou* (denotes English blue and green). In Experiment 1, English children who had not yet marked a boundary linguistically responded categorically. The Himba children in the current experiment had not yet acquired their color terms; Himba children acquire color terms at around 5 or 6 years of age (Androulaki, 2003). Therefore, one aim of the current experiment was to test whether Himba children (who have no color term knowledge) would respond categorically. A second aim of the current experiment was to compare the extent of the category effect for the Himba toddlers with that for the English toddlers.

The universalistic and perceptual reorganization models would predict that the Himba toddlers will respond categorically across the blue–purple boundary. The universalistic model would predict that the extent of the category effect will be equal for English and Himba children. The perceptual reorganization model would predict that because color term knowledge and linguistic categorization is more advanced for the English, the category effect will be stronger for the English children than for the Himba children. The linguistic relativity model would predict that the Himba toddlers will not respond categorically at all.

Method

Participants

The Himba keep no record of birth dates; therefore, the ages of children had to be estimated. All children were probably under 5 years of age given that no children could touch their ear with their hand while reaching across their head. Children under 5 years of age do not pass this "ear-touching" test due to the size of the head relative to the limbs at that age (Gabriel, 2001). Other physical characteristics (e.g., height) suggested that no children were under 2 years of age. Additional information was obtained from discussions with the children's mothers using landmark events, such as Namibian independence, to establish a time frame. The mean age of the children as a group was estimated to be 3 years. Consequently, the ages of the English and Himba sample were not matched exactly, although both groups consisted of toddlers. In Experiment 1, 60 English toddlers were tested. In Experiment 2, because of the difficulties in conducting research in remote rural areas, the potential sample size of toddlers was small. Therefore, there were 32 children in the sample (18 boys and 14 girls).

Stimuli, design, and procedure

A version of the same task as in Experiment 1 was used; however, because the Himba do not have teddy bears, a cardboard cutout of a Himba boy was used instead of a bear. Himba children who had not yet learned color terms were tested. A quick check on naming on a small sample of the children (using the comprehension and naming tasks) confirmed that the children tested had not yet learned color terms; therefore, naming and comprehension data were not collected subsequently. Because only 32 Himba toddlers were available to take part in Experiment 2, it was decided that for this experiment only one boundary (blue–purple) would be tested. The blue–purple boundary was chosen because the Himba adults occasionally use borrowed terms to mark the blue–green and pink–red boundaries, yet there is no borrowed term for purple. The stimuli were identical to the blue–purple stimuli used in the 2-AFC task in Experiment 1. The task was completed outside, but not in direct sunlight or deep shade (color temperature = 5500–7000 K, as indicated by a Gossen colormaster 3F). The task was conducted by the second author, but instructions were given to the children through a Namibian translator. Children were tested individually where they were free from distraction.

Results

2-AFC accuracy scores

The Himba were more accurate for the between-category pair (mean accuracy = 68.75%, $SD = 23.76$) than for the within-category pair (mean accuracy = 56.25%, $SD = 22.89$), $t(31) = 2.56$, $p < .05$. Accuracy was significantly above chance for the between-category pair, $t(31) = 4.46$, $p < .001$, but not for the within-category pair, $t(31) = 1.54$, $p = .13$.

Comparison of the Himba mean accuracy scores with the English scores for the blue-purple boundary from Experiment 1 shows that although the English children performed better than the Himba children overall, the size of the category effect did not differ for the Himba and English children. This was supported by an ANOVA looking at the effects of language (Himba or English) and category (between or within). Category was a repeated measures factor. Accuracy was higher for English toddlers (mean = 77.50%, $SD = 21.30$) than for Himba toddlers (mean = 62.50%, $SD = 18.78$), $F(1, 50) = 7.08$, $MSE = 782.5$, $p < .05$. Participants were more accurate for the between-category pair (mean = 75.96%, $SD = 23.20$) than for the within-category pair (mean = 60.57%, $SD = 27.27$), $F(1, 50) = 15.91$, $MSE = 407.5$, $p < .001$. There was no significant interaction between category and language, $F(1, 50) = 0.85$, $MSE = 407.5$, $p = .36$.

Discussion

The Himba toddlers showed a significant category effect across the blue-purple boundary. Performance was at chance for within-category pairs and was approximately 70% for between-category pairs. Therefore, toddlers with no implicit or explicit color term knowledge at all showed categorical perception on a 2-AFC task. The Himba participants were less accurate than the English participants. This might be due to various factors such as familiarity with formalized memory tasks. There was no significant difference in the extent of the category effect for Himba and English toddlers. As with Experiment 1, the results of Experiment 2 provide support for the universalistic model. However, can we be sure that the categorical responding shown in Experiments 1 and 2 is actually categorical perception? Experiment 3 addressed this issue.

Experiment 3: The nature of the category effect in toddlers

The findings of Experiments 1 and 2 suggest that language may have a minimal role in the origin and early development of the category effect. However, it is possible that the effect reflects memory factors rather than categorical perception. If this were the case, then one would expect poorer categorical perception performance on tasks with high memory demands and would expect better categorical perception performance on tasks with low memory demands. This hypothesis was tested in Experiment 3.

Experiment 3 investigated the nature of the category effect further by manipulating the length of the delay on the 2-AFC task. If the category effect is a memory effect, then when the memory component of the task is reduced (no delay condition) or removed (simultaneous condition), the category effect should weaken or disappear. If the category effect is perceptual, then the category effect should remain even when the memory component is removed (simultaneous condition).

Method

Participants

A total of 59 English-speaking children (29 boys and 30 girls) between 24 and 47 months of age took part in the experiment (mean age = 36.2 months, *SD* = 6.5). There were 20 children in each condition (delay condition mean age = 36.5 months, *SD* = 6.2; no delay condition mean age = 36.3 months, *SD* = 6.8; simultaneous condition mean age = 35.8 months, *SD* = 6.7).

Stimuli, design, and procedure

The stimuli were identical to the pink–red stimuli used in the 2-AFC task in Experiment 1. Ideally, all boundaries would be tested; however, because of the large sample size that would be needed for this (approximately 180 toddlers), this was not practical. Therefore, the pink–red set was focused on due to the novel nature of the boundary; this boundary is defined by differences in lightness and saturation rather than hue.

There were three different task procedures. The delay procedure was the same as in Experiments 1 and 2; there was a 5-s delay before stimulus choice. For the no delay procedure, there was no delay before stimulus choice. Therefore, the target was presented on Bear A for 5 s, Bear A and Bear A's jumpers were covered, and immediately afterward the target and foil for Bear B were presented. For the simultaneous procedure, the target jumper on Bear A was shown simultaneously with the target and foil jumpers of Bear B.

Results

2-AFC accuracy scores

Fig. 6 gives the mean accuracy scores for within- and between-category pairs under the three conditions. It appears that the length of delay has no effect on the strength of the category effect.

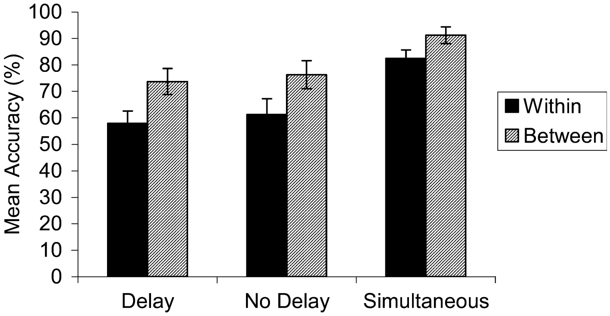

Fig. 6. Mean accuracy (±1 *SE*) for within- and between-category pairs for the 5-s delay, no delay, and simultaneous conditions.

An ANOVA looking at the effects of task condition (5-s delay, no delay, or simultaneous) and category (between or within) on the accuracy scores supports this. Category was a repeated measures factor. There was a main effect of task condition on accuracy: 5-s delay (mean = 65.8%, $SD = 16.6$), no delay (mean = 68.7%, $SD = 20.5$), and simultaneous (mean = 86.9%, $SD = 11.8$), $F(2, 56) = 9.26$, $MSE = 556$, $p < .001$. Bonferroni post hoc tests revealed significant differences in accuracy between the simultaneous condition and the no delay condition, $t(58) = 3.44$, $p < .005$, and between the simultaneous condition and the 5-s delay condition, $t(58) = 3.95$, $p < .005$, but not between the no delay condition and the simultaneous condition, $t(58) = 0.55$, $p = 1$. As in Experiments 1 and 2, participants were more accurate for between-category pairs (mean = 80.4%, $SD = 21.3$) than for within-category pairs (mean = 67.2%, $SD = 23.3$), $F(1, 56) = 17.7$, $MSE = 289$, $p < .001$. Importantly, there was no significant interaction between the task condition and category, $F(2, 56) = 0.51$, $MSE = 289$, $p = .60$. A t test confirmed that a category effect was found for the simultaneous condition, $t(19) = 2.33$, $p < .05$.

Discussion

If the category effect was a memory effect, then the size of the effect should decrease as the length of the delay was reduced and no category effect should be found when the memory component of the task was removed altogether. Experiment 3 found a category effect when there was no memory component to the task. In addition, the length of the delay on the 2-AFC task had no impact on the strength of the categorical responding. This is consistent with the category effect being due to a perceptual process; it seems as though the category effect is indeed categorical perception. The implications of this, and the findings of Experiments 1 and 2, are discussed further in the General discussion.

General discussion

Summary of the main findings

Categorical perception was shown by toddlers across a range of boundaries on a 2-AFC task. The extent of categorical perception was not affected by linguistic categorization, and the strength of the category effect did not increase with increased general color term knowledge. Himba toddlers who had acquired no color terms showed categorical perception across the blue–purple linguistic boundary—a boundary that is not marked linguistically by their parental language. There was no significant difference in the extent of the category effect for Himba and English toddlers. In Experiment 3, the size of the category effect was not affected by the length of the delay on the 2-AFC task, and categorical responding was shown even when there was no memory component to the task. The results of Experiment 3 suggest that the category effect found in Experiments 1 and 2 is likely to be truly perceptual. There are various other inferences we can make from this set of results. These are discussed in turn.

Categorical perception in the absence of language

The infant studies (Bornstein et al., 1976; Franklin & Davies, 2004) revealed that categorical responding to color can be shown in the absence of language. However, because habituation and novelty preference tasks were used in these studies, it is not entirely clear whether these category effects are equivalent to the category effects found in adult studies of categorical perception. The current experiments provide evidence of category effects in toddlers with no color term knowledge at all, using a 2-AFC task, as in the adult studies of categorical perception. Therefore, these results add robustness to the claim that the origin of categorical perception is not language. The results provide a bridge between the infant and adult studies of categorical perception.

Color term acquisition and categorical perception

The cross-cultural studies (e.g., Roberson et al., 2000) could suggest that language has an effect on categorical perception. In the current experiments, using a method similar to that of Roberson and colleagues (2000), we found that at a certain developmental level, there was no evidence that language affects categorical perception. Moreover, because the extent of categorical perception does not vary with linguistic categorization in English children, and the extent of categorical perception does not differ between English and Himba children, categorical perception appears not to be strengthened by color term acquisition. The findings of the current experiments suggest that language has a minimal role in the origin and early development of categorical perception.

Universality of prelinguistic categorical perception

As discussed in the Introduction, it is theoretically possible for cross-cultural differences to exist before language acquisition. The current experiments found no evidence of cross-cultural differences in categorical perception between Himba and English toddlers despite cultural differences in visual environment. Therefore, the findings of the current experiments suggest that prelinguistic categorical perception may be universal.

Implications for the debate

In summary, the evidence from the current experiments overwhelmingly supports universalistic models of categorical perception; no evidence has been found to support linguistic relativity or perceptual reorganization models of categorical perception. However, how can we account for the cross-cultural differences in adult studies of categorical perception? One possibility is that these cross-cultural differences in adults' color categorization are not truly perceptual and are actually a result of verbal labeling (Munnich & Landau, 2003). The argument here is that if nonlinguistic perceptual tasks are used cross-culturally, then cross-cultural differences in categorical perception would not be shown and the underlying perceptual categorization of color would be revealed to be universal. It would, of course, be interesting to test for categorical perception in Himba adults across the blue–purple boundary and other boundaries using such nonlinguistic perceptual tasks.

A second possibility is that perceptual reorganization does occur but that language may only modify categorical perception later in life—after color terms have become cognitively ingrained. Consistent with the argument that it may take time for language to reorganize perception, Lucy and Gaskins (2001, 2003) showed that preferences for classification by shape or material are universal at 7 years of age yet do not become language specific until approximately 9 years of age. For perceptual reorganization of color, it might take extensive repeated labeling for language to amplify the existing category effect, and language-specific effects might not arise until later during childhood. Cross-cultural longitudinal studies of categorical perception are needed to assess this. Such studies would also allow a test of other aspects of the perceptual reorganization model, for example, whether categorical perception shown during infancy is later lost if a language does not mark the boundary and whether categorical perception not shown during infancy may later be acquired if a language does mark the distinction between the two colors.

At first glance, it appears that there is abundant evidence that the influence of language on perception and cognition is pervasive, occurring in many domains. These include time (e.g., Boroditsky, 2001), space (e.g., Bowerman & Choi, 2001; Majid, Bowerman, Kita, Haun, & Levinson, 2004; McDonough, Choi, & Mandler, 2003), and object perception (Boroditsky, Schmidt, & Phillips, 2003) (see also Gentner & Goldin-Meadow, 2001). It is tempting, perhaps, to let this pervasiveness influence our belief in language affecting color perception. Caution is required, however, for two reasons. First, color could be an exception. Boroditsky and colleagues (2003) argued that language may be more powerful in influencing domains such as time than in influencing color due to time being more abstract and less reliant on sensory experience than color. But more crucially, the evidence for pervasiveness might not be as clear-cut as it initially seems. The concerns about whether linguistic effects on categorical perception of color are due to labeling also apply to other domains because some of the studies in other domains also use tasks that can be completed using a labeling strategy (Hermer-Vazquez, Spelke, & Katsnelson, 1999; Jackendoff & Landau, 1991; Malt, Sloman, & Gennari, 2003).

This investigation has raised and discussed the possibility that universal hardwired color categories are molded and reorganized by linguistic input during development. As outlined in the Introduction, similar models have been proposed by Werker and colleagues (Werker & Lalonde, 1988; Werker & Tees, 1983, 1984) for the development of speech perception and by Hespos and Spelke (2004) for conceptual development. It is also possible that universal hardwiring may be molded by linguistic input in other areas of cognitive and perceptual development, for example, the perception of time. However, the impact of language need not necessarily be the same for all domains. As Hespos and Spelke stated, "Intuition suggests a difference between mature auditory and conceptual capabilities.

The effects of language experience therefore may be more dramatic at the interface of audition and phonology than at the interface of conceptual structure and semantics'' (p. 455). Further research into the impact of language learning on cognitive and perceptual development is needed to address these issues.

Further questions and the wider issues

The argument that there is an innate predisposition for category boundaries at certain locations in the color space raises many questions. For example, why does the innate location of perceptual category boundaries align with English color terms rather than the color terms of other languages such as the Himba language? Why do languages vary in their color terminology if the structure of perceptual categorization is initially universal? What causes language to ignore the innate perceptual organization? For example, how is it possible that a linguistic category (e.g., *grue*, a term describing both green and blue) spans the blue–green perceptual boundary? Dedrick (1997) argued, ''We need to reconsider and reconceive the path that will take us from innate perceptual saliencies to basic (and perhaps other) color language'' (p. 187). He added, ''We should consider there to be a space between the perceptual and the linguistic which needs to be filled by an account of the rules that people use to generate relatively stable reference classes in a social context'' (p. 187). The combination of psychological, anthropological, philosophical, physiological, and linguistic approaches to the issue of color categorization may provide more answers.

Conclusions

The experiments presented here have provided evidence that language is not the origin of categorical perception and that language does not modify categorical perception in toddlers. Therefore, these experiments provide further support for universalistic theories of perceptual color categorization. No evidence was provided to support the relativistic claim that language is the origin of the categorical perception effect. No evidence was provided to support the theory that language modifies or modulates the location and extent of categorical perception as color terms are acquired. It remains possible that cross-cultural differences in adults are not truly perceptual or that language perceptually reorganizes categorical perception later in development. Further research is needed to explore these possibilities and to further explain the apparent inconsistencies in the debate about the origin and nature of color categorization.

Acknowledgments

This research was supported by an Economic and Social Research Council postgraduate studentship (R42200124191) and an Economic and Social Research Council postdoctoral fellowship (PTA-026-27-0110) to the first author. We owe thanks to

Amy Riddett for assistance with some of the data collection in Namibia and to Keemu Jakarama for translation. We are very grateful to the toddlers who participated in this research. We are also grateful to Gavin Bremner, Paul Sowden, Christine Daoutis, Samantha Boyles, and four reviewers for their constructive comments.

References

Andrick, G. R., & Tager-Flusberg, H. (1986). The acquisition of color terms. *Journal of Child Language, 13*, 119–134.

Androulaki, A. (2003). *Colour term acquisition and the development of working memory in children: A cross-linguistic investigation and a test of the linguistic relativity hypothesis*. Unpublished PhD thesis, University of Surrey, Guildford, UK.

Atkinson, J. (1984). Human visual development over the first 6 months of life: A review and a hypothesis. *Human Neurobiology, 3*, 61–74.

Bloom, P. (2004). Children think before they speak. *Nature, 430*, 410–411.

Bornstein, M. H. (1973). Color vision and color naming: A psychophysiological hypothesis of cultural difference. *Psychological Bulletin, 80*, 257–285.

Bornstein, M. H. (1985). On the development of color naming in young children: Data and theory. *Brain and Language, 26*, 72–93.

Bornstein, M. H., Kessen, W., & Weiskopf, S. (1976). Color vision and hue categorization in young infants. *Journal of Experimental Psychology: Human Perception and Performance, 1*, 115–129.

Bornstein, M. H., & Korda, N. (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. (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. 61–81). Cambridge, MA: 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. 475–512). Cambridge, MA: Cambridge University Press.

Daoutis, C. A., Franklin, A., Riddett, A., Clifford, A., & Davies, I. R. L. (2004). *Category effects in children's colour search: A cross-linguistic comparison.* Unpublished manuscript.

Daoutis, C. A., Pilling, M., & Davies, I. R. L. (2004). *Categorical effects in visual search for colour*. Unpublished manuscript.

Davies, I., & Franklin, A. (2002). Categorical similarity may affect color pop-out in infants after all. *British Journal of Developmental Psychology, 20*, 185–203.

Davies, I. R. L., Özgen, E., Pilling, M., & Riddett, A. (2004). The relationship between categorical perception and memory bias towards the prototype. *Perception, 33*, 66.

Dedrick, D. (1997). Color categorisation and the space between perception and language. *Behavioral and Brain Sciences, 20*, 187–188.

Fahle, M., & Poggio, T. (2002). *Perceptual learning*. Cambridge, MA: MIT Press.

Franklin, A., & Davies, I. R. L. (2004). New evidence for infant color categories. *British Journal of Developmental Psychology, 22*, 349–378.

Franklin, A., Pilling, M., & Davies, I. R. L. (2004). Category effects in visual search for colour: Evidence from eye-movement latencies. *Perception, 33*, 147.

Gabriel, D. J. (2001). *The spirit of childhood*. Cambridge, UK: RSC Press.

Gegenfurtner, K. R., & Kiper, D. C. (2003). Color vision. *Annual Review of Neuroscience, 26*, 181–206.

Gentner, D., & Goldin-Meadow, S. (Eds.). (2001). *Language in mind*. Cambridge, MA: MIT Press.

Glucksberg, S., Hay, A., & Danks, J. H. (1976). Words in utterance contexts: Young children do not confuse the meaning of same and different. *Child Development, 47*, 737–741.

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

Goldstone, R. L., Lippa, Y., & Shiffrin, R. M. (2001). Altering object representations through category learning. *Cognition, 78*, 27–43.

Guenther, F. H., Husain, F. T., Cohen, M. A., & Shinn-Cunningham, B. G. (1999). Effects of categorisation and discrimination training on auditory perceptual space. *Journal of the Acoustical Society of America, 106*, 2900–2912.

Harnad, S. (1987). Psychophysical and cognitive aspects of categorical perception: A critical overview. In S. Harnad (Ed.), *Categorical perception: The groundwork of cognition* (pp. 287–301). New York: Cambridge University Press.

Hermer-Vazquez, L., Spelke, E. S., & Katsnelson, A. S. (1999). Sources of flexibility in human cognition: Dual-task studies of space and language. *Cognitive Psychology, 39*, 3–36.

Hespos, S. J., & Spelke, E. S. (2004). Precursors to spatial language. *Nature, 430*, 453–456.

Huttenlocher, J., Hedges, L. V., & Vevea, J. L. (2000). Why do categories affect stimulus judgment?. *Journal of Experimental Psychology: General, 129*, 220–241.

Jackendoff, R., & Landau, B. (1991). Spatial language and spatial cognition. In D. J. Napoli & J. A. Kegl (Eds.), *Bridges between psychology and linguistics: A Swarthmore festschrift for Lila Gleitman* (pp. 145–169). Hillsdale NJ: Lawrence Erlbaum.

Jameson, K. A., Highnote, S., & Wasserman, L. (2001). Richer color experience for observers with multiple photopigment opsin genes. *Psychonomic Bulletin & Review, 8*, 244–261.

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

Livingston, K. R., Andrews, J. K., & Harnad, S. (1998). Categorical perception effects induced by category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 732–753.

Lucy, J. A., & Gaskins, S. (2001). Grammatical categories and the development of classification preferences: A comparative approach. In M. Bowerman & S. C. Levinson (Eds.), *Language acquisition and conceptual development* (pp. 257–284). Cambridge, UK: Cambridge University Press.

Lucy, J. A., & Gaskins, S. (2003). Interaction of language type and referent type in the development of nonverbal classification preferences. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind* (pp. 465–493). Cambridge, MA: MIT Press.

MacLeod, D. I. A. (2003). Colour discrimination, colour constancy, and natural scene statistics. In J. D. Mollon, J. Pokorny, & K. Knoblauch (Eds.), *Normal and defective colour vision* (pp. 189–217). Oxford, UK: Oxford University Press.

Majid, A., Bowerman, M., Kita, S., Haun, D. B. M., & Levinson, S. C. (2004). Can language restructure cognition? The case for space. *Trends in Cognitive Sciences, 8*, 108–114.

Malan, J. S. (1995). *Peoples of Namibia*. Pretoria, South Africa: Rhino.

Malt, B. C., Sloman, S. A., & Gennari, S. P. (2003). Speaking versus thinking about objects and actions. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind* (pp. 81–113). Cambridge, MA: MIT Press.

McDonough, L., Choi, S., & Mandler, J. M. (2003). Understanding spatial relations: Flexible infants, lexical adults. *Cognitive Psychology, 46*, 229–259.

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* (pp. 113–157). Cambridge, MA: MIT Press.

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.

Özgen, E., & Davies, I. R. L. (2003). *Acquisition of categorical colour perception: New evidence suggests perceptual change*. Paper presented at the meeting of the European Society for Cognitive Psychology.

Petzold, A., & Sharpe, L. T. (1998). Hue memory and discrimination in young children. *Vision Research, 38*, 3759–3772.

Pilling, M., & Davies, I. R. L. (in press). Linguistic relativity and colour cognition. *British Journal of Psychology*.

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

Pitchford, N. J., & Mullen, K. T. (2002). Is the acquisition of basic-color terms in young children constrained?. *Perception, 31*, 1349–1370.

Prinzmetal, W., Amiri, H., Edwards, T., & Allen, K. (1998). Phenomenology of attention: I. Color, location, orientation, and spatial frequency. *Journal of Experimental Psychology: Human Perception and Performance, 24*, 261–282.

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

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

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

Sapir, E. (1921). *Language*. New York: Harcourt, Brace and World.

Shatz, M., Behrend, D., Gelman, S. A., & Ebeling, K. S. (1996). Color term knowledge in two-year-olds: Evidence for early competence. *Journal of Child Language, 23*, 177–199.

Shepard, R. N. (1992). The perceptual organization of colors: An adaptation to the regularities of the terrestrial world. In J. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind* (pp. 494–532). Oxford, UK: Oxford University Press.

Sugita, Y. (2004). Experience in early infancy is indispensable for color perception. *Current Biology, 14*, 1267–1271.

Uchikawa, K., & Shinoda, H. (1996). Influence of basic color categories on color memory discrimination. *Color Research and Application, 21*, 430–439.

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.

Werker, J. F., & Lalonde, C. E. (1988). Cross-language speech perception: Initial capabilities and developmental change. *Developmental Psychology, 24*, 672–683.

Werker, J. F., & Tees, R. C. (1983). Developmental changes across childhood in the perception of non-native speech sounds. *Canadian Journal of Psychology, 37*, 278–286.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development, 7*, 49–63.

Whorf, B. L. (1956). In J. B. Caroll (Ed.), *Language, thought, and reality: Essays by B. L. Whorf.* Cambridge, MA: MIT Press.

Yendrikhovskij, S. N. (2001). Computing color categories from statistics of natural images. *Journal of Imaging Science Technology, 45*, 409–417.