

GOLDSMITHS Research Online
Article

Goldstein, Julie, Davidoff, Jules B. and Roberson, Debi

Knowing color terms enhances recognition: Further evidence from English and Himba

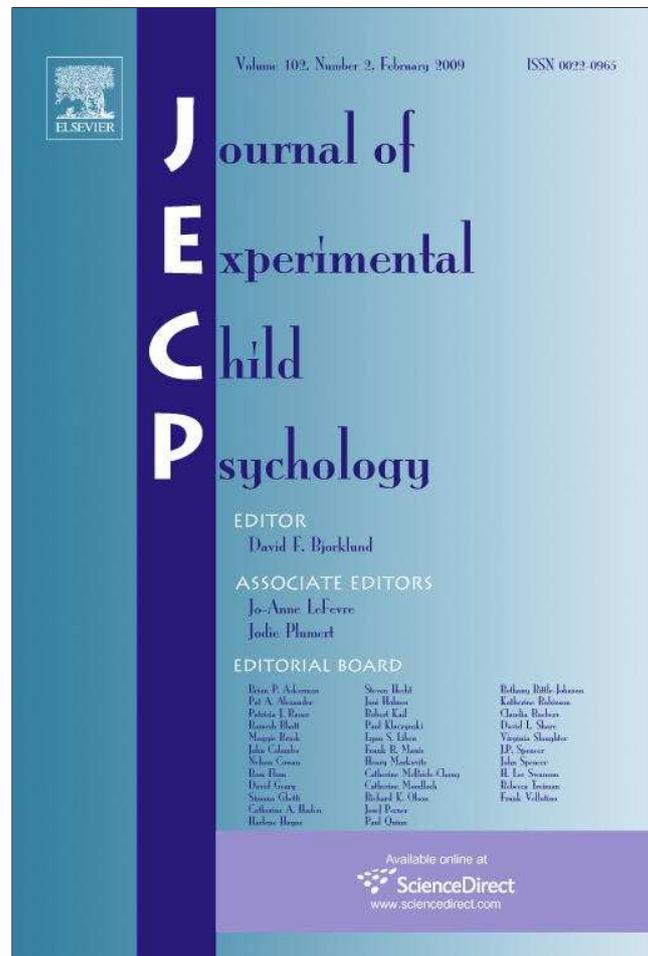
You may cite this version as: Goldstein, Julie, Davidoff, Jules B. and Roberson, Debi. 2009. Knowing color terms enhances recognition: Further evidence from English and Himba. *Journal of Experimental Child Psychology*, 102(2), pp. 219-238. ISSN 00220965 [Article] : Goldsmiths Research Online.

Available at: <http://research.gold.ac.uk/4935/>

COPYRIGHT

All material supplied via Goldsmiths Library and Goldsmiths Research Online (GRO) is protected by copyright and other intellectual property rights. You may use this copy for personal study or research, or for educational purposes, as defined by UK copyright law. Other specific conditions may apply to individual items.

This copy has been supplied on the understanding that it is copyright material. Duplication or sale of all or part of any of the GRO Data Collections is not permitted, and no quotation or excerpt from the work may be published without the prior written consent of the copyright holder/s.

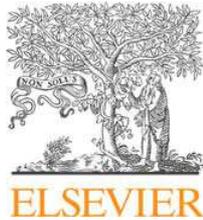


This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>



Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

journal homepage: www.elsevier.com/locate/jecp

Knowing color terms enhances recognition: Further evidence from English and Himba

Julie Goldstein^a, Jules Davidoff^{a,*}, Debi Roberson^b^a Department of Psychology, Goldsmiths, University of London, London SE14 6NW, UK^b Department of Psychology, University of Essex, Colchester CO4 3SQ, UK

ARTICLE INFO

Article history:

Received 29 February 2008

Revised 17 June 2008

Available online 8 August 2008

Keywords:

Color term acquisition

Categorical perception

ABSTRACT

Two experiments attempted to reconcile discrepant recent findings relating to children's color naming and categorization. In a replication of Franklin and colleagues (*Journal of Experimental Child Psychology*, 90 (2005) 114–141), Experiment 1 tested English toddlers' naming and memory for blue–green and blue–purple colors. It also found advantages for between-category presentations that could be interpreted as support for universal color categories. However, a different definition of knowing color terms led to quite different conclusions in line with the Whorfian view of Roberson and colleagues (*Journal of Experimental Psychology: General*, 133 (2004) 554–571). Categorical perception in recognition memory was now found only for children with a fuller understanding of the relevant terms. It was concluded that color naming can both underestimate and overestimate toddlers' knowledge of color terms. Experiment 2 replicated the between-category recognition superiority found in Himba children by Franklin and colleagues for the blue–purple range. But Himba children, whose language does not have separate terms for green and blue, did not show a cross-category advantage for that set; rather, they behaved like English children who did not know their color terms.

© 2008 Published by Elsevier Inc.

Introduction

The proposal put forward by Berlin and Kay (1969) for universal color categories has not been entirely abandoned by its adherents (Kay & Regier, 2003; Regier, Kay, & Cook, 2005), but the results of

* Corresponding author.

E-mail address: j.davidoff@gold.ac.uk (J. Davidoff).

recent studies of color language and cognition have promoted some compromise between that nativist view and a relativist (often called Whorfian) view for the origin of color categories (Drivonikou et al., 2007; Franklin et al., 2008; Gilbert, Regier, Kay, & Ivry, 2006; Kay & Regier, 2006; Pitchford & Mullen, 2005; Tan et al., 2008). These studies have shown, in different ways and under different conditions that cognitive organization of color corresponds at least in some part to the color terms of the speaker's language.

The allowance that color language contributes to the establishment of color categories has been necessary to accommodate those recent studies that showed a tight relationship between color language and color cognition across different languages and cultures (Davidoff, Davies, & Roberson, 1999; Roberson, Davidoff, Davies, & Shapiro, 2004; Roberson, Davidoff, Davies, & Shapiro, 2005; Roberson, Davies, & Davidoff, 2000; Roberson, Pak, & Hanley, 2008; Winawer, Witthoft, Frank, Wu, & Boroditsky, 2007). These studies show no evidence that color "space" is warped only at the boundaries between a putative set of hardwired and innate color categories as proposed by Kay and McDaniel (1978) and Rosch (1973). Indeed, a recent functional magnetic resonance imaging (fMRI) study by Tan et al. (2008) suggested that language processing areas in the brain are directly involved even in the simplest discriminations of easy-to-name colors.

In the studies of Davidoff, Roberson, and colleagues, shades of color that are equally spaced in a particular perceptual metric (e.g., the Munsell system) were judged as unequally spaced by speakers of a language that marks a category boundary between the colors. In other words, the enhanced discrimination of colors from different categories over equally separated colors within the same category, generally termed *categorical perception* (see Harnad, 1987), has been produced by the color terms of the speaker's language. For example, adult speakers of Berinmo and Himba, neither of which marks the distinction between blue and green, show no evidence of better discrimination and memory at the blue–green boundary but do show those effects at the boundary between color categories that they mark linguistically, whereas English does not (Roberson et al., 2000; Roberson et al., 2005). Similar behavioral differences have been reported between English speakers and speakers of Russian (Winawer et al., 2007) and Korean (Roberson et al., 2008), both of which have more color terms than English.

Despite converging evidence for the importance of color terms in establishing categorical perception for adult populations, there is a considerable body of evidence showing it to be present as early as 4 months of age (Bornstein, Kessen, & Weiskopf, 1976; Franklin & Davies, 2004; Franklin, Pilling, & Davies, 2005b; Franklin et al., 2008). Indeed, in Franklin and Davies (2004), categorical perception was found in 4-month-olds between secondary colors as well as between primary colors, in places where universalist theories (e.g., Kay & Regier, 2003) would not predict it. There is a contrary finding to Gerhardstein, Renner, and Rovee-Collier (1999), but Davies and Franklin (2002) contended that the article's conclusions are unreliable due to unintended category boundaries. For a strong relativist position (Davidoff, 2001), the infant data are hard to accommodate. If color category divisions were established through language, one would not expect any evidence for categorical perception before the acquisition of color terms. However, the diversity of color terms found in the world's languages must mean that any proposed innate category divisions can be overruled by language-induced categories (Bornstein, 1985; Franklin et al., 2008). The current article does not address the infant data, but clearly at some point researchers will need to reconcile them with the adult data that show color categorical perception to be language dependent (Gilbert et al., 2006; Roberson et al., 2000; for recent fMRI evidence of language processing in the discrimination of easy-to-name color stimuli, see also Tan et al., 2008).

The current article addresses the evidence from toddlers where there is evidence in favor of both nativist and relativist positions. Children acquire color terms in no particular order (Heider, 1972; Roberson et al., 2004) except for the tardiness of gray and brown (Pitchford & Mullen, 2002; Pitchford & Mullen, 2003; Pitchford & Mullen, 2005). The critical issue here is whether that acquisition coincides with the onset of categorical perception. In the first developmental cross-cultural study, Roberson et al. (2004) showed a longitudinal shift in recognition errors consistent with the development of categorical perception. Franklin, Clifford, Williamson, and Davies (2005a) reported differently, arguing that there is a particular set of panhuman color categories that are established not by language but rather by the properties of human color vision. Their argument

was based on data from samples from both a Western population (United Kingdom) and the same remote population (the Himba of Namibia) used in Roberson et al. (2004). In Franklin et al. (2005a), U.K. 2- and 3-year-olds showed better discrimination of colors that crossed the boundary between blue and green (as well as between blue and purple) than of colors that were fully within either category (the hallmark of categorical perception) irrespective of their color naming ability. Furthermore, the Himba 2- to 5-year-olds, who of course could not be expected to know English color terms, nevertheless showed evidence of categorical perception at the boundary between blue and purple. Evidence more in line with Roberson et al. (2004) comes from Daoutis, Franklin, Riddett, Clifford, and Davies (2006). Using a visual search paradigm in a study of children from the United Kingdom, Himba, and another similar language group, they found that search times were quicker for colors embedded within colors of a different category only for children who gave those categories different names. Therefore, it seemed appropriate to replicate Franklin et al. (2005a) because their data must create problems of interpretation for relativist positions.

For the replication of their U.K. toddler study, we followed closely the procedures of Franklin et al. (2005a) but used only their blue–purple and blue–green ranges and did not include the color range pink–red because we did not want to introduce the complication of both hue and lightness differences in determining a color category. In addition, we wanted to be more certain as to whether or not children comprehended the relevant color terms. Franklin and colleagues, carried out three analyses. The first was an assessment of categorical perception in their whole sample, the second was an analysis of categorical perception with respect to correct naming of the stimuli used in the two-alternative forced-choice task, and the third was an examination of the role of color naming fluency (combined naming and comprehension of 11 focal examples of basic colors) on the categorical perception of blue–green and blue–purple. The measure of comprehension used was a child's ability to correctly select a color from the 11 focal examples when asked (“Show me a red one”). In Franklin and colleagues' study, a small negative relationship was found between color fluency and the extent of categorical perception. The authors offered no explanation of this puzzling finding, and this is another reason for attempting to replicate their study. Critically, we added a fourth analysis in which we took into account not only the occasions on which a term was used correctly (“hits”) but also those on which it was used incorrectly (“false alarms”). The fourth analysis addresses our concern that the tests in Franklin and colleagues' study may have underestimated the color knowledge of children.

For their Himba investigation of categorical perception, Franklin et al. (2005a) chose to investigate only the blue–purple boundary. The blue–purple range seems to be an unusual choice given that it may contain a boundary for Himba even though the Himba language does not have a purple category as such. The majority of adult Himba speakers extend their dark term *zoozu* to a range of stimuli that an English speaker would call *purple* (see Appendix A), and most would apply their blue–green term *burou* to the range of blue stimuli used by Franklin and colleagues. So, if the Himba children had attained some understanding of their color terms, it may have contributed toward the finding of categorical perception. Therefore, in repeating their investigation, we also added an examination of a range of colors around the blue–green boundary for which adult Himba speakers would use only the single term *burou*.

Experiment 1: English children

Method

Participants

A total of 60 native English-speaking children (37 girls and 23 boys) between 31 and 51 months of age (mean age = 42.24 months, $SD = 5.87$) were recruited from four different nurseries and a music group for toddlers in the London area. None of the children was color-blind, according to information provided by the schools.

Stimuli

Naming and comprehension. Stimuli were sweater shapes cut out of colored paper that could be placed over cardboard bear figures. The 22 colors used for naming and comprehension tasks (see Roberson et al., 2004) were made from Color-Aid paper. These colors included the 11 best examples of the basic English color terms (red, green, blue, yellow, pink, purple, orange, and brown) used by Franklin et al. (2005a), hereafter termed “focal colors” (Berlin & Kay, 1969). An additional set of 11 nonfocal poor exemplars were added, as in Roberson et al. (2004), and were chosen to be intermediate between the basic chromatic categories of English (e.g., halfway between pink and orange). The focal colors were as follows (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). The nonfocal colors were as follows: blue (BV HUE), pink–purple (RVR HUE), yellow (YOY HUE), orange (O HUE), green (GBG HUE), pink–red (ORO T2), green (YGY T1), orange (OYO T4), orange–yellow (YOY T3), purple (RVRT3), and pink (ROR T2). Appendix B gives the designations and CIE L*u*v* coordinates for each color.

Two-alternative forced-choice discrimination task. For this task, the colored sweaters were made from glossy Munsell colored paper. There were two sets of test stimuli, blue–green and blue–purple, with three stimuli per set. The Munsell notations of the stimuli for the blue–green set were 7.5G, 5BG, and 2.5B, all with value (lightness) 5 and chroma (saturation) 10. The stimuli for the blue–purple set were 10B, 7.5PB, and 5P, all with value 3 and chroma 10. Thus, the test stimuli in each set varied only in Munsell hue, with lightness and saturation kept constant. The test stimuli in each set are referred to as A1, A2, and B, where A1 and A2 belong to one category and B belongs to a different category, so as to analyze naming patterns at test (as in Franklin et al., 2005a).

For each of the blue–green and blue–purple sets, there was a within-category pair (A1, A2) and a between-category pair (A2, B) for which the separations of within- and between-category pairs were equated in Munsell units of perceptual distance (see Fig. 1). There were three Munsell steps between stimuli in a pair. Although the human color space on which the Munsell set of colors is based is non-Euclidian, and completely uniform color differences are impossible to represent in a

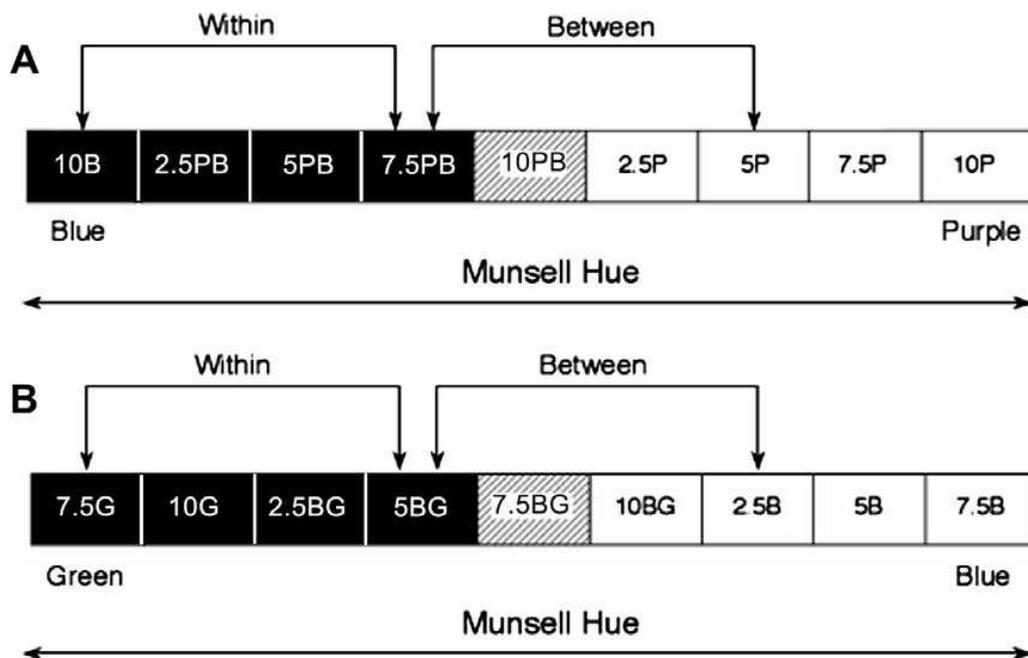


Fig. 1. Munsell codes, categorical status, and Munsell distances of the stimuli of the category pairs used in Experiment 1. The categorical relationships (within or between) of the category pairs are shown for blue–purple (A) and green–blue (B). In Panel A, chroma = 10 and value = 3. In Panel B, chroma = 10 and value = 5. (Figure taken from Franklin et al., 2005a).

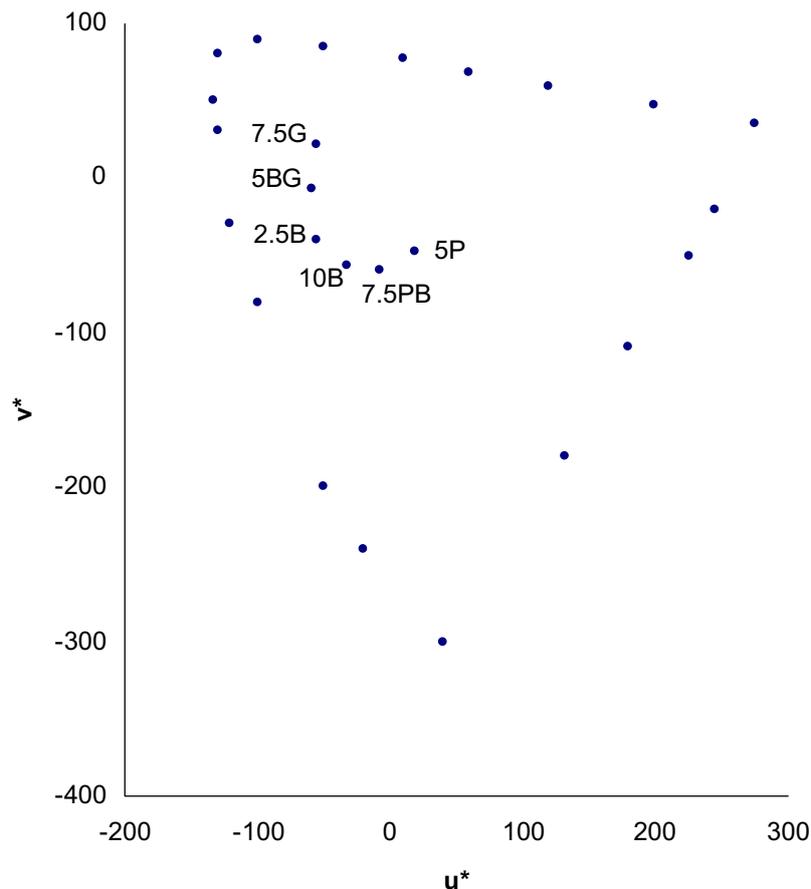


Fig. 2. Munsell colors of Fig. 1 represented in CIE $L^*u^*v^*$ color space.

three-dimensional color model (MacEvoy, 2005), Franklin et al. (2005a) equated color differences in the more objective CIE $L^*u^*v^*$ perceptual space (see Fig. 2). They concluded that test stimuli and foils were equally separated in color space, lending further support to the equidistance between colors suggested by Munsell (1905). The location of the category boundaries was taken from the adult data in Franklin and Davies (2004). The Munsell codes and categorical status of stimulus pairs are shown in Fig. 1A and B (from Franklin et al., 2005a). Blue–green and blue–purple separations were approximately 30 units in CIE color space ($30 \Delta E$).

Procedure

Overall design. All testing was conducted under standardized lighting conditions with the use of a Gretag Macbeth lamp (D65, 6500 K, at 810–1880 lux), as in Franklin et al. (2005a), with children seated at a comfortable distance from stimuli to point to them. All children completed the training, naming, and comprehension task using the 22 colored sweaters. For the two-alternative forced-choice task, children were tested on both sets of stimuli, in contrast to Franklin and colleagues' study, where different children were tested on the two sets. Half of the children received the blue–purple set first, and the other half received the blue–green set first. For each continuum, eight category pairs were randomly presented until the four between- and four within-category pairs were exhausted. The two-alternative forced-choice task was always presented after the naming task and comprehension tasks that were presented in a randomized order.

Comprehension and naming of colors. For the comprehension task, the 22 focal and nonfocal sweaters were laid out in front of the child, and the child was asked to point to a named color (e.g., “Can you show me a red one?”). The color of the chosen sweater was recorded, and the experimenter then asked the child whether there were any other sweaters that corresponded to this

specific color (e.g., “Are there any other red ones?”). The child’s selection was noted by the experimenter, who then asked, “Is there another red one?” The question was repeated, and additional answers were recorded until the child said that there were no more sweaters of that color. The same comprehension procedure was carried out for all 11 focal color names in random order. For naming, each of the 22 sweaters was presented individually and in random order, and the child was asked to name the color of the sweater. The procedure here differs from that used by Franklin et al. (2005a), who assessed naming and comprehension using only the 11 focal colors. In their study, children were not asked whether there were other colors that could be given the same name.

Two-alternative forced-choice task. A training session was carried out before the two-alternative forced-choice task as in Franklin et al. (2005a). The aim of the training session was to show each child that when Bear A wore a particular sweater, Bear B also wore an identically colored sweater. Because the meanings of *same* and *different* may be poorly understood by young children (Glucksberg, Hay, & Danks, 1976), the term *different* was not used in the procedure and the training phase ensured that the child understood the meaning of the word *same* in the following manner. The two 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 an identical set of 11 focal sweaters was 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. Next, 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 was uncovered and the child was asked to find the same sweater for Bear B from 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 the criterion of three successive correct responses was reached. All of the children reached criterion.

In the two-alternative forced-choice task, the sweaters were presented under standardized lighting conditions that simulate natural daylight using the Gretag Macbeth lamp to maintain the uniformity of the Munsell system (Davies & Franklin, 2002). 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. 1A and B). The procedure was repeated four times for each within- and between-category stimulus pair (eight trials in all). For two of each set of four trials, one stimulus was the target and the other stimulus was the foil; for the remaining two trials, this target/foil allocation was reversed. The order of presentation was randomized for each child. Alternate children carried out the testing first with the blue–purple set or blue–green set. After the experimental trials, the child was presented with each of the stimuli individually and was asked to name the color of the sweater.

Results

Four analyses were carried out on the English children’s data: three analyses identical to Franklin et al. (2005a) and one additional analysis. The first analysis considered, for all children, category effects for the blue–purple and blue–green sets in the two-alternative forced-choice task. The second analysis addressed the effects of naming ability for the two-alternative forced-choice colors on the category effect in the first analysis. The third analysis addressed color term fluency, assessed with focal colors, in relation to the categorical effects of the first analysis. The fourth (additional) analysis was a variation of the first analysis and assessed whether children do or do not know

color terms from the naming of focal colors and the false alarms made in the comprehension of the color words.

Analysis 1: Categorical perception for blue–purple and blue–green

For each child, the number of correct identifications in the two-alternative forced-choice task was calculated when the choice was between two stimuli from the same category (within-category pairs) and when the choice was between stimuli from 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 for within- and between-category pairs for the blue–purple and blue–green sets. As in Franklin et al. (2005a), between-category accuracy was higher than within-category accuracy for all conditions.

Accuracy scores were investigated in a 2 (Task Order: blue–purple set first vs. blue–green set first) \times 2 (Category: between vs. within) \times 2 (Set: blue–purple vs. blue–green) mixed-design analysis of variance (ANOVA) with repeated measures over the last two variables. The only significant effect was the main effect of category. Accuracy was greater for between-category pairs than for within-category pairs, $F(1, 58) = 59.48$, $p < .001$, $h_p^2 = .306$. No other effects approached significance, all $F_s < 1$, except for the Category \times Set interaction, $F(1, 58) = 1.04$, $p > .30$, $h_p^2 = .018$. Accuracy was above chance (2/4) for both the blue–purple set, $t(59) = 9.94$, $p < .001$, and the blue–green set, $t(59) = 13.21$, $p < .001$, and a Bonferroni post hoc test found no significant difference in accuracy between sets, $t(59) < 1$. As in Franklin et al. (2005a), categorical perception was demonstrated equally for both sets.

Analysis 2: Effect of naming accuracy on the size of the category effect

Table 1 gives the naming frequencies for each of the two-alternative forced-choice stimuli for both sets. Overall, the majority name is between 59 and 85% in agreement with the adult linguistic boundary for the blue–purple and blue–green sets. Of the three stimuli, A2 seems to be most ambiguous for both sets, probably because it is closer to the linguistic boundary than are A1 and B. These dips are also found in Franklin et al. (2005a) Table 1, although to a lesser extent.

Individual naming patterns for the colors used in the two-alternative forced-choice task were analyzed to investigate the effect of linguistic categorization on categorical perception in the manner of Franklin et al. (2005a). 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 group*). A “don’t know” response was also counted as a term. Therefore, two patterns of naming qualified: (A1: name 1; A2: name 1; B: name 2) and (A1: name 1; A2: name 1; B: “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

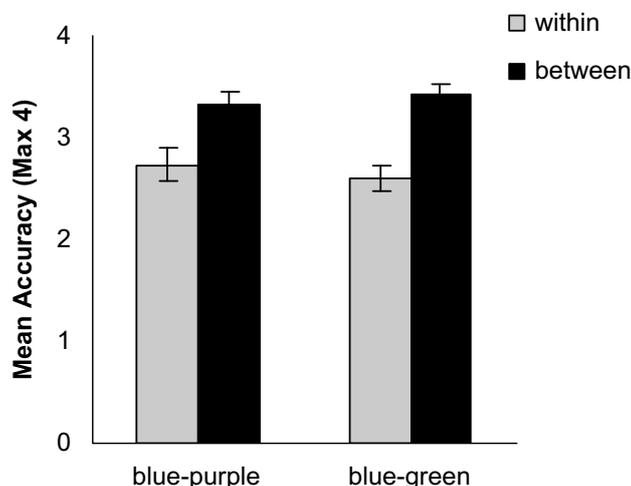


Fig. 3. Mean accuracy (± 1 SE) for all children for within- and between-category pairs for the blue–purple and blue–green sets.

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	83.6
		Purple	6.8
		Green	4.9
		Red	4.9
		Blue	59
	A2: 7.5PB 3/10	Purple	31.2
		Green	6.6
		Red	1.6
		Don't know	1.6
		Purple	83.6
	B: 5P3/10	Blue	6.6
		Red	3.3
		Yellow	3.3
		Green	1.6
		Pink	1.6
Blue–green	A1: 7.5G 5/10	Green	85.3
		Blue	3.3
		Red	3.3
		Yellow	3.3
		Purple	1.6
		Pink	1.6
	A2: 5BG 5/10	Orange	1.6
		Green	65.6
		Blue	21.3
		Red	6.6
		Don't know	3.3
		Purple	1.6
	B: 2.5B 5/10	Yellow	1.6
		Blue	85.2
		Green	6.6
		Pink	3.3
		Red	3.3
		Yellow	1.6

linguistic boundary (*reverse boundary* group). Again, a “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: “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 (*no boundary* group). Again, “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: “don't know”; A2: “don't know”; B: “don't know”).

Franklin et al. (2005a) combined naming accuracy across the different sets because their sample sizes were not sufficient to investigate the effect of language for each boundary separately. The current study, by using a within-participants design, was able to investigate effects of language for each boundary separately. Thus, participants were grouped into a name boundary, reverse boundary, or no boundary group for each set separately.

For the blue–purple set, 30 children had a name boundary, 10 had no name boundary, and 20 had a reverse boundary. For the blue–green set, 38 children had a name boundary, 14 had no name boundary, and 8 had a reverse boundary. In Franklin et al. (2005a), children scoring at floor or ceiling were not removed from the analysis. Given their prediction of categorical perception for all groups, inclusion of those children would only add noise to their data and, if anything, strengthen any conclusions if categorical perception were found. However, in the current context that predicts no difference between within- and between-category conditions, inclusion of those children might promote a false null result. So, in all subsequent analyses, data are presented only after exclusion of children performing at chance or at ceiling.

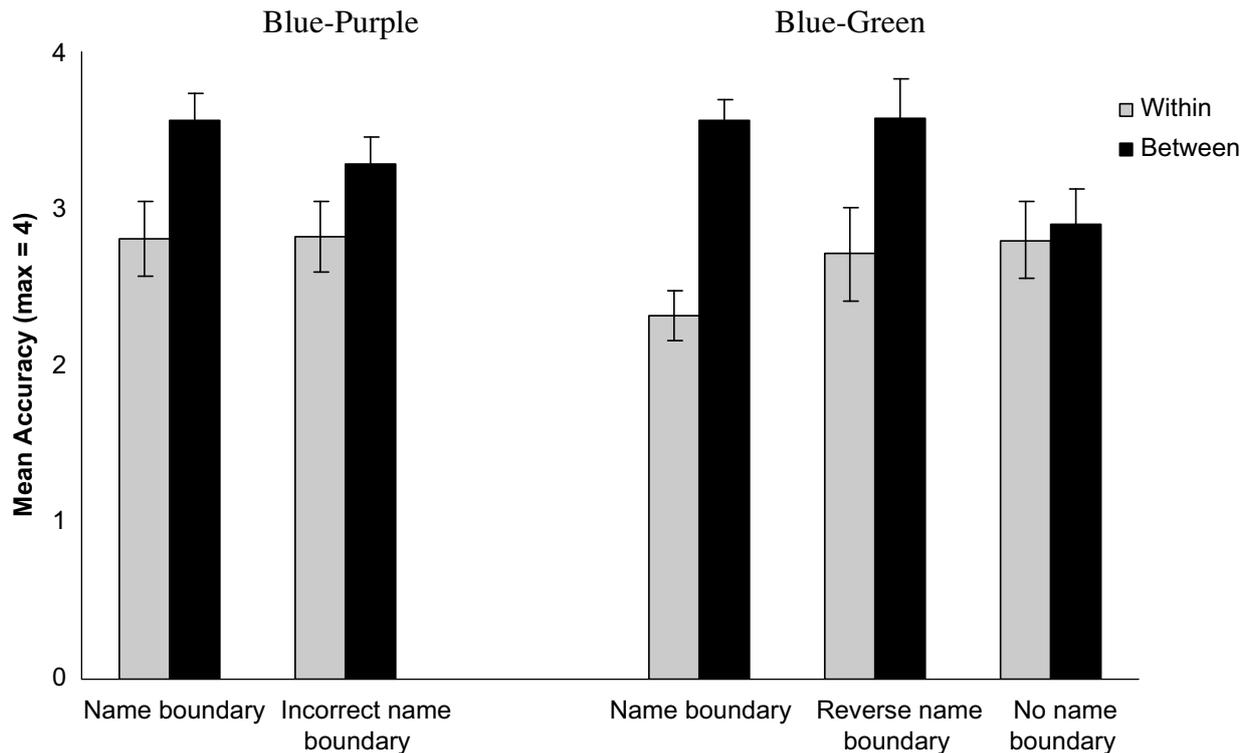


Fig. 4. Mean accuracy (± 1 SE) of children depending on their individual name boundary status for the blue–purple and blue–green sets on within- and between-category pairs. For blue–purple, name boundary ($n = 16$) and incorrect name boundary ($n = 17$); for blue–green, name boundary ($n = 25$) and reverse name boundary ($n = 7$), No name boundary ($n = 10$).

After exclusion of children at floor or ceiling for the blue–purple set, 16 children had a name boundary, 14 had a reverse boundary, and 3 had no name boundary. Therefore, analysis was undertaken comparing those children who had the correct name boundary with those who did not. Accuracy was analyzed in a 2 (Name Boundary: name boundary vs. incorrect boundary) \times 2 (Category: within vs. between) mixed-design ANOVA. As in Franklin et al. (2005a), the only significant effect was the main effect of category, $F(1, 31) = 5.93$, $p = .022$, $h_p^2 = .158$, with better performance on between-category trials overall (see Fig. 4). The main effect of name boundary was not significant, $F(1, 31) < 1$. The Category \times Name Boundary interaction was also not significant, $F(1, 31)$. After exclusion of children at floor or ceiling for the blue–green set, 25 children had a name boundary, 7 had a reverse boundary, and 10 did not have a name boundary. Accuracy was analyzed in a 3 (Name Boundary: boundary vs. no boundary vs. reverse boundary) \times 2 (Category: within vs. between) mixed-design ANOVA with repeated measures over the second factor. There was a significant main effect of category, $F(1, 39) = 10.04$, $p = .003$, $h_p^2 = .205$, with no main effect of name boundary, $F(2, 39) = 1.40$, $p < .30$, $h_p^2 = .067$, and a trend for the Category \times Name Boundary interaction, $F(2, 39) = 2.74$, $p = .077$, $h_p^2 = .123$. Bonferroni adjusted post hoc tests revealed that the difference between within- and between-category pairs was significant only for the name boundary group, $t(24) = 4.89$, $p < .001$. The reverse boundary group escaped significance, $t(6) = 2.12$, $p = .078$, and the no boundary group showed no significant difference, $t(9) < 1$. Here, with the no boundary group (see Fig. 4), we have an indication of a contrast to the findings of Franklin et al. (2005a) of equal categorical perception for all groups.

Analysis 3: Color term fluency and categorical perception

Franklin et al. (2005a) analyzed color term acquisition (number of focal colors named and comprehended) in 2- to 4-year-olds. These children were split into four age groups covering 6 months each (2–2.5, 2.5–3, 3–3.5, and 3.5–4 years). We were unable to find sufficient children between 2 and 2.5 years who could successfully carry out our tasks, so participants' age in our study spanned

Table 2

Percentages of (a) correctly named focal colors in the naming task and (b) correctly identified focals in the comprehension task for each age group as assessed by the procedures used in Franklin et al. (2005a)

Color	2.5–3 years (<i>n</i> = 12)		3–3.5 years (<i>n</i> = 23)		3.5–4 years (<i>n</i> = 17)		4–4.5 years (<i>n</i> = 8)		Total (<i>N</i> = 60)	
	a	b	a	b	a	b	a	b	a	b
Black	58.3	66.7	60.9	78.3	88.2	88.2	88.9	88.9	74.08	80.53
White	33.3	75.0	82.6	82.6	88.2	76.4	88.9	77.8	73.25	77.95
Red	75.0	66.7	69.5	73.9	94.1	64.7	88.9	77.8	81.88	70.78
Green	50.0	58.3	78.3	65.2	94.1	76.5	88.9	66.7	77.83	66.68
Yellow	75.0	66.7	82.6	78.3	88.2	94.1	100.0	77.8	86.45	79.23
Blue	58.3	58.3	78.3	78.3	100.0	94.1	100.0	77.8	84.15	77.13
Orange	58.3	66.6	60.9	60.9	88.2	82.3	77.7	55.6	71.25	66.35
Pink	75.0	75.0	73.9	65.2	88.2	88.2	88.9	77.8	81.50	76.55
Purple	58.3	58.3	78.3	73.9	82.3	76.5	88.9	77.8	76.95	71.63
Brown	9.1	41.7	43.4	47.8	64.7	58.8	66.6	55.5	45.95	50.95
Gray	33.3	25.0	26.1	69.6	41.2	64.7	55.6	66.6	39.05	56.48
Total	53.08	59.85	67.58	70.36	83.40	78.59	84.85	72.74	72.42	70.38

from 2.5 to 4.5 years. In consequence, our four age ranges were a little different. Our groups were aged 2.5 to 3 years (*n* = 12), 3 to 3.5 years (*n* = 23), 3.5 to 4 years (*n* = 17), and 4 to 4.5 years (*n* = 8). Table 2 gives the percentages of correct responses for each focal color as assessed in Franklin and colleagues' study.

Children's naming and comprehension of the 11 focal colors improved from a mean of approximately 6 terms for the youngest age group to a mean of approximately 9.5 terms for the oldest age group (see Table 2); these are slightly different from the rates reported by Franklin et al. (2005a), who found somewhat better naming and comprehension for children corresponding to our two youngest age ranges. In the current study, as in Franklin and colleagues' study, there was no difference between naming and comprehension rates, $t(10) < 1$ (mean naming = 72.42%, mean comprehension = 70.38%). Using similar naming and comprehension tasks, some researchers have found that young children pass naming tests before they comprehend color terms (e.g., Soja, 1994).

The relationship between color term fluency (focal colors named and comprehended) and the extent of the category effect was explored here in the same way as in Franklin et al. (2005a). Two indexes were calculated: a color term fluency index, averaging the mean number of focal colors named and identified, and a categorical effect index, subtracting the within-category score from the between-category score (on the two-alternative forced-choice 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.

For the blue–purple set, the relation between the color term fluency and categorical effect indexes was not significant, $r = -.13$, $p > .20$. For the blue–green set, the relation between color term fluency and categorical effect indexes also was not significant, $r = .12$, $p > .30$. Unlike in Franklin et al. (2005a), where there was an unexplained small negative correlation, here there appeared to be no effect of color term fluency on categorical perception.

Analysis 4: Categorical perception with a revised assessment of the understanding of color terms

Pitchford and Mullen (2003) discussed how best to assess a toddler's comprehension of color terms. They argued, following Soja (1994), that it is insufficient to simply examine hits and misses when the child is asked to point to a color. In their analysis, they were able to use a signal detection analysis to take into account false alarms. Their rigorous procedure required too many trials for our younger children, so we assigned children as knowing colors following Soja's procedure previously used with English toddlers by Pitchford and Mullen (2001) and also by Roberson et al. (2004) with both English and Himba toddlers. Franklin et al. (2005a) assessed comprehension by asking children to point to a color

from among only the focal colors (the best examples of each category). Here we added nonfocal colors to the array to make the subsequent request, “Are there any more [color] ones?,” a sensible one for the child. The presence of poor examples as well as focal colors gave a more realistic test of the child’s category knowledge.

In the current study, for the blue–purple set, children were defined as knowing those color terms if they named focal examples of blue and purple and pointed to these colors in the comprehension test correctly and no more than once when probed to name other colors. A similar definition applied for the blue–green set. If a child pointed to a nonfocal color that could be given the same name, then of course this was not regarded as an error. However, if a child pointed at a focal color more than once when probed for another name (e.g., pointing to the blue tile when asked to find both green and purple tiles), then the child was considered as not knowing this term. Thus, children who did not know the color terms were defined as those who did not give the correct names for focal colors of blue and purple or blue and green or who pointed to these colors on more than one occasion when probed for other colors.

Table 3 shows the classification of the children, after the exclusion of those at floor or ceiling, under the criteria used in Franklin et al. (2005a) and also according to the Soja (1994) criteria. It is clear that our criteria changed the allocation. A very similar pattern would have emerged if all 60 children had been included in Table 3. In Franklin et al. (2005a), children were allocated according to their performance on the stimuli used in the two-alternative forced-choice task. The Soja (1994) procedure allocated children according to their ability to name focal colors and not make false alarms when asked to point to those colors. For the most part, fewer children were classified as knowing color terms with Soja’s procedure. For example, with the blue–green set, 25 children were assigned to the linguistic boundary group using the Franklin et al. (2005a) naming criteria, but only 17 children were assigned to the linguistic boundary group by the Soja (1994) criteria, thereby leaving 8 children with an incomplete understanding of blue and green. However, allocation to groups is of equal concern for the no boundary group, where some children clearly did have some comprehension of colors. Of the 13 children classified in the no boundary group by their naming in the two-alternative forced-choice task, 4 actually knew the relevant color terms. Thus, using the Soja’s criteria, some children in the current study were reallocated as either having or not having a linguistic boundary.

In sum, there were 30 children who knew all of the blue, green, and purple terms and 12 children who did not know any of these terms. In addition, there were 7 children who knew only blue and purple, 4 children who knew only blue and green, 2 children who knew only green, and 1 child who knew only purple. In total, 37 children knew blue and purple, and 14 children knew neither of these terms; also, 34 children knew blue and green, and 13 children knew neither of these terms. It is only these children who were allocated to the “know” and “don’t know” groups; children who knew one term but not the other of the two terms in a set were excluded.

Table 3

Numbers of children, excluding those at floor or ceiling, who were allocated to linguistic boundary, reverse linguistic boundary, and no boundary groups according to Franklin et al. (2005a) criteria of correctly naming the colors used in the two-alternative forced-choice task and their knowledge of color terms as assessed by the criteria of Soja (1994) for either blue–purple or blue–green

Set	Boundary group	Allocation by the procedures in Franklin et al. (2005a)	Children knowing color terms by the procedures in Soja (1994)
Blue–purple	Linguistic boundary	16	14
	Reverse linguistic	14	8
	No boundary	3	1
Blue–green	Linguistic boundary	25	17
	Reverse linguistic	7	3
	No boundary	10	3

Note. The Soja (1994) criteria consist of correctly naming focal colors and not making false alarms when asked to point to the colors that match a spoken color name.

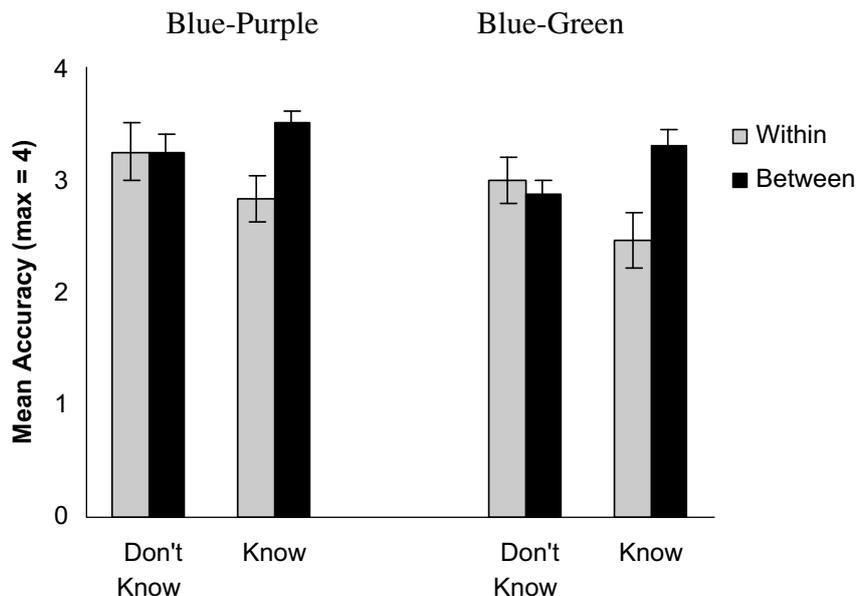


Fig. 5. Mean accuracy (± 1 SE) by color term knowledge (“don’t know” vs. “know”) for within- and between-category pairs for blue–purple and blue–green sets, respectively. For both sets, “don’t know” ($n = 8$) and “know” ($n = 23$).

To ensure that any null effects were not exaggerated by the inclusion of floor and ceiling effects, our comparison of “know” and “don’t know” children, as in Analysis 2, excluded those children. For the blue–purple set, this left 23 children in the “know” group and 8 in the “don’t know” group. For the blue–green set, this also left 23 children in the “know” group and 8 in the “don’t know” group.

We found that only when children knew their color terms did they show evidence of categorical perception. This conclusion was supported by two 2 (Category: between vs. within) \times 2 (Linguistic Terminology: know vs. don’t know) ANOVAs with repeated measures over the first factor. Fig. 5 shows the mean accuracy on the two-alternative forced-choice task for those children who either knew or did not know both terms in a particular set.

For the blue–purple set, the main effect of category, $F(1, 29) = 4.45, p = .041, h_p^2 = .128$, was significant, with better performance on between-category trials. However, the significant interaction, $F(1, 29) = 4.45, p = .041, h_p^2 = .128$, indicated that only children who knew their color terms showed a significantly better performance on between-category trials, $t(22) = 3.89, p = .001$; don’t know: $t(7) < 1$ (see Fig. 5). The main effect of linguistic terminology was not significant, $F(1, 29) < 1$, showing no difference in performance across terminology groups.

For the blue–green set, there was also a main effect of category, $F(1, 29) = 4.56, p = .042, h_p^2 = .136$, with better performance on between-category trials. The significant interaction, $F(1, 29) = 7.38, p = .011, h_p^2 = .203$, again indicated categorical perception for children who knew the relevant terminology, $t(22) = 4.70, p < .001$, but not for those who had not acquired those color terms, $t(7) < 1$ (see Fig. 5). The main effect of linguistic terminology was not significant, $F(1, 28) < 1$. Thus, even when ceiling and chance performances were eliminated from the analyses, the results indicated that when color term knowledge (i.e., naming and comprehension) were taken into account, categorical perception was found only for children who knew their color terms.

Discussion

The current study set out to replicate and extend Franklin et al. (2005a). As in that study, an analysis of accuracy on the two-alternative forced-choice task (Analysis 1), without regard for age, color term knowledge, or floor and ceiling effects, showed categorical perception effects across both the blue–purple and blue–green boundaries. A second analysis, based on naming accuracy for the colors

used in the two-alternative forced-choice task, gave results that mostly replicated Franklin and colleagues' findings. However, for the blue–green set, we found that children with no name boundary did not show categorical perception. The third analysis, concerning children's color term knowledge by age group, showed that they had acquired between 6 and 9.5 terms across the age span tested, slightly lower than the levels reported by Franklin and colleagues. Another difference was that, in Franklin and colleagues' study, color term knowledge was significantly negatively correlated—although not a large effect—with the extent of categorical perception. The current study found no relationship between category fluency and categorical perception. However, it was really only Analysis 4 that gave data with an emphasis different from those found in Franklin and colleagues' study. In Analysis 4, where children were divided by whether they could correctly name focal colors and comprehend the range for which the color terms were appropriate, categorical perception was restricted to those who knew both color terms in a set.

Analysis 4 leads to an important conclusion with respect to the role of color naming in categorical perception. The fact that children do not correctly name the colors used in the two-alternative forced-choice task does not necessarily imply a complete lack of understanding of color categories. Children may show incorrect color naming for marginal exemplars of a category but still have some comprehension of the category (for how sharpening of a color category can occur by exclusion of colors from distant categories prior to exclusion of colors from adjacent categories, see also Pitchford & Mullen, 2003). Therefore, it is possible that, in Franklin et al. (2005a), the inclusion of children in the no boundary group (Analysis 2) may have contributed to the color categorical perception for that group. Conversely, children may correctly name marginal examples but still not see colors from the same category as more similar than colors from different categories. Thus, it is not simply the online use of color terms that produces color categorical perception (Fonteneau & Davidoff, 2007; Franklin et al., 2008).

Experiment 2: Himba toddlers

In Experiment 2, we replicated and extended the cross-lingual aspect of Franklin et al. (2005a) study with Himba toddlers who were unlikely to have acquired their own color terms (Androulaki, 2003, cited in Franklin et al., 2005a). Franklin and colleagues found categorical perception for Himba toddlers who, whether or not they had their own color terms, clearly would not have had the English terms *blue* and *purple* used in that study. Their result fits with a nativist account rather than a relativist account. As discussed earlier, the decision to use blue–purple may have created an ambiguous situation with regard to Himba color terms. Franklin and colleagues justified not using the more obvious blue–green continuum because the Himba term *burou*, covering the range of colors between blue and green, was a borrowed color term. However, this concern seems misplaced because that term (*burou*) would have been used by adults for some of the stimuli in their blue–purple continuum. So, in addition to replicating their study, we also carried out the test with a continuum between the primary colors of blue and green.

Method

Participants

Testing took place on two field trips to Namibia. As closely as allowed, the first Himba group of 36 toddlers (21 boys and 15 girls ranging between 2 and 4.5 years of age), tested during the first field trip, and the second Himba group of 24 toddlers (10 boys and 14 girls also ranging between 2 and 4.5 years of age), tested during the second field trip, were matched in age to the group of 60 English toddlers from Experiment 1. As in Franklin et al. (2005a), the age of the Himba toddlers needed to be estimated because the Himba do not keep track of their age in the same way as Westerners do. Contralateral ear touching (Gabriel, 2001), physical height, and any other information the parents could give us (e.g., age in relation to sibling) were used to estimate the age of each child.

Stimuli, design, and procedure

Stimuli, design, and procedure followed Experiment 1 very closely, although bears were replaced by cut-out figures of boys as in Franklin et al. (2005a). Naming and comprehension data were not collected by Franklin and colleagues because, from an initial check, the children were found not to know color terms. The same applies to the current data. The same blue–purple set was chosen to replicate Franklin and colleagues' findings, but also the blue–green set from Experiment 1 was used in this within-participants design. As in Franklin and colleagues' study, for the first Himba group of children, testing was carried out in shaded natural daylight under a tree with color temperature assessed using a light meter. With the second Himba group, a lightbox was used to control for any illumination variation that might have affected the appearance of the colored shirts. The lightbox used Illuminant C (a bluish light simulating natural daylight) under which the Munsell samples are standardized. A Namibian translator helped with instructions. Each child was tested individually, free from distraction.

Results

As with the English children, for each child, the number of correct identifications on the two-alternative forced-choice task was calculated when the choice was between two stimuli from the same category (within-category pairs) and when the choice was between stimuli from different categories (between-category pairs). The maximum number of correct identifications was four for within-category pairs and four for between-category pairs. The Himba children carried out the task for both color sets balanced for order. In the critical analyses of category, children who were at floor or ceiling on the task were excluded. To make sure that order effects were unimportant for that analysis, the accuracy data were first analyzed summing over the two sets in a 2 (Task Order: blue–purple set first vs. blue–green set first) \times 2 (Category: between vs. within) \times 2 (Lighting: natural daylight vs. lightbox) mixed-design ANOVA with repeated measures over the factor category. No effects were significant, all p s $>$.13.

For the blue–purple set, 17 children were at ceiling and 4 were at or below chance, leaving 39 children for the analysis of category effects. The data from the 39 children were analyzed in a 2 (Category: between vs. within) \times 2 (Lighting: natural daylight vs. lightbox) mixed-design ANOVA with repeated measures over the first factor. The only significant effect was the main effect of category, $F(1,$

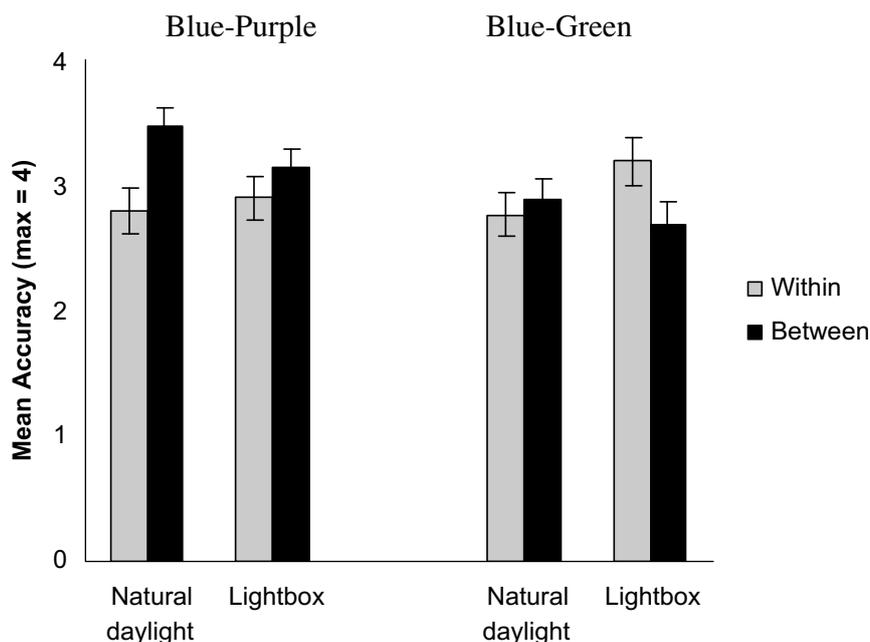


Fig. 6. Himba mean accuracy (± 1 SE) for within- and between-category trials for blue–purple (natural daylight, $n = 19$; lightbox, $n = 20$) and blue–green (natural daylight, $n = 17$; lightbox, $n = 16$) sets.

37) = 7.71, $p < .011$, $h_p^2 = .162$, with better performance on between-category trials (see Fig. 6). There was no main effect of lighting, $F(1, 37) < 1$, and no significant interaction between lighting and category, $F(1, 37) = 1.55$, $p < .30$, $h_p^2 = .04$.

For the blue–green set, 27 children at or below chance were removed, and the remaining 33 children were submitted to a similar analysis. Neither main effect was significant, $F(1, 31) < 1$, and the interaction was also not significant, $F(1, 31) = 2.45$, $p < .20$, $h_p^2 = .073$ (see Fig. 6).

Discussion

Our results replicate those of Franklin et al. (2005a) for the blue–purple set, whereas they give a quite different outcome for the blue–green set. The performance of Himba children in the blue–green set resembled that of the “don’t know” group of English toddlers, with no evidence of categorical perception. Importantly, because a relativist view of color cognition would predict a null result, the toddlers were well above chance on the task. Therefore, results of the blue–green set further support the view that acquisition of color terms is required to show categorical perception.

Both Franklin et al. (2005a) and the current study showed categorical perception for Himba toddlers using the blue–purple set of colors. We wondered whether testing under natural daylight conditions in Franklin and colleagues’ study could have produced the category effect for the blue–purple set, but this seems unlikely given that performance under controlled lighting was essentially similar to that replicating the conditions in their study. We would rather suggest that category effects for the blue–purple set arise because of some knowledge of the Himba color terms *zozu* and *burou* (see Appendix A) or because they are better acquainted with cattle terms than with color terms (Goldstein & Davidoff, 2008). In considering the Himba names for colors, Roberson et al. (2004), Roberson et al. (2005) found extensive use of terms describing animal skin patterns (combinations of color, pattern, and texture) for several colors and confirmed that the preferred term used by adults for colors in the blue–purple range was not always a color term. Indeed, for the three colors used in the blue–purple continuum in Experiment 2, the purple color was named equally often with a cattle term as with the dominant color term. However, the two blue colors were called by the dominant color term approximately four times more frequently. Thus, if cattle terms were known by the children, they would likely have been used only for the purple color and, hence, set up an artifactual boundary.

General discussion

A replication of Franklin et al. (2005a) study was carried out with additional analyses that might help to examine the role of toddlers’ color naming and comprehension in color categorical perception. Despite some small procedural differences in the replication from the original study (within-group rather than between-group, naming undertaken before rather than after the two-alternative forced-choice task, and exclusion of the youngest age group), the results were initially very similar to those of Franklin and colleagues. If anything, one might have predicted that the procedural changes would encourage categorical perception in our “don’t know” group of children. For example, it might have been thought that prior testing of focal color naming and comprehension before the two-alternative forced-choice task would encourage color naming for the stimuli used in the two-alternative forced-choice task and artifactually produce the between-category superiority in the two-alternative forced-choice task (Munnich & Landau, 2003), but the data show otherwise. Indeed, despite the remarkable findings in Roberson and Davidoff (2000), now replicated in many laboratories, that perceptual similarity (categorical perception) changes with verbal load, we know from the procedures introduced in Franklin et al. (2005b) that overt naming is not the origin of categorical perception.

Experiment 1 found an overall effect of category for both color sets for the sample of 60 children (Analysis 1). However, with such a large age range, a different performance by a relatively small number of children who did not know color terms might not be revealed in the group analysis. Hence, as in Franklin et al. (2005a), children were allocated according to their color naming ability for the stimuli used in the two-alternative forced-choice task (Analysis 2). Using their procedure, we showed substantial agreement with Franklin and colleagues' data for the blue–purple range, although we did not fully replicate their findings of equivalent categorical perception independent of the accuracy of color naming for the blue–green range. More striking is the finding that a different assessment of children's understanding of color terms produced a quite different outcome. Critical here is how we define the understanding of a color term. The procedure adopted by Franklin and colleagues of dividing children by their naming the colors used in the two-alternative forced-choice task may appear to be a reasonable way to assess the knowledge of a color term. However, we argue that for young children it is more important to have some estimate of the range of colors that children find acceptable for the color term. Only then can we be confident that colors from within the same category are perceived as being more similar to each other than would be assumed from their separation in $L^*u^*v^*$ space. More rigorously, one might use a signal detection procedure (Pitchford & Mullen, 2002; Pitchford & Mullen, 2003), but the simpler procedure from Soja (1994) was sufficient to cause a rather different interpretation of the data. In the current study, color term acquisition was reanalyzed using a method that split children into “know” and “don't know” groups (Analysis 4). Now the only group that showed category effects was the group of children who knew their color names according to Soja's criteria that take into account false alarms as well as accurate naming of focal colors.

Within our Analysis 2, we replicated Franklin et al. (2005a) intriguing finding that a substantial number of children would be classified as having a reverse boundary because their name boundary between green and blue (and also between purple and blue) was not yet aligned to that of adults. Franklin and colleagues made little comment on this reverse boundary group. Clearly, both their data and our data found that categorical perception does not simply align with actual naming, but there is more than one possible interpretation. It might be interpreted to show that color term acquisition is not necessary for the establishment of categorical perception. However, we suggest that the names reported by the children do not properly reflect their underlying color knowledge (see also Pitchford & Mullen, 2003; Roberson et al., 2004). Table 3 shows that nearly half of the children classified as having a reverse boundary should actually be classified as knowing the color terms, supporting the second interpretation.

With respect to Analysis 3, Franklin et al. (2005a) suggested that if color term acquisition were needed to establish categorical perception, then there could be a positive relation between color term fluency and categorical perception. This argument presumes that a relativist account would predict a positive correlation between the extent of categorical perception for any particular color and the number of other color terms known by children. In fact, they found an unexplained small, but nevertheless significant, negative correlation. However, there is no reason to assume a relationship between categorical perception for a particular boundary and the total number of terms known. When children definitely know a color term, this implies that they will not confuse that color with other nearby terms but says nothing about their knowledge of other regions of color space. There is no reason to suppose that acquiring more secondary terms (e.g., knowing orange and pink) would make a significant difference for categorical perception to green, blue, or purple, and the current data found no such correlation.

Considering the Himba data, Experiment 2 replicated the cross-category advantage found by Franklin et al. (2005a) for the blue–purple range of colors but did not find the same for the additional blue–green range of colors. Yet the boundary between blue and green is one that is still regarded as potentially universal (Kay & Regier, 2003), whereas that between blue and purple is not. The lack of categorical perception for the blue–green range is not due to the Himba children's misunderstanding the task given that their performance was well above chance. We further examined whether testing under natural daylight conditions in Franklin et al. (2005a) could explain the differences between the two sets of results, but this seems unlikely because

the pattern of results under controlled lighting was essentially similar to that replicating the conditions in their study. Above, we proposed that it is the use of noncolor words for the purple color this is the most probable origin of a boundary between blue and purple for the Himba toddlers.

In summary, although our results at first appeared to be very similar to those of Franklin et al. (2005a), a more detailed analysis does not support their conclusion that there is one particular set of universal color categories (those named in English) that are independent of language acquisition (Bornstein et al., 1976; Franklin & Davies, 2004). The counterargument here points to the importance of color term acquisition by children of Himba and English cultures in establishing both color categories and categorical perception. These findings reflect differences across cultures in accord with other color categorization studies (Davidoff et al., 1999; Roberson et al., 2004; Roberson et al., 2005; Roberson et al., 2008; Winawer et al., 2007) and among toddlers of the same culture separated by language acquisition (O'Hanlon and Roberson, 2006; Roberson et al., 2004).

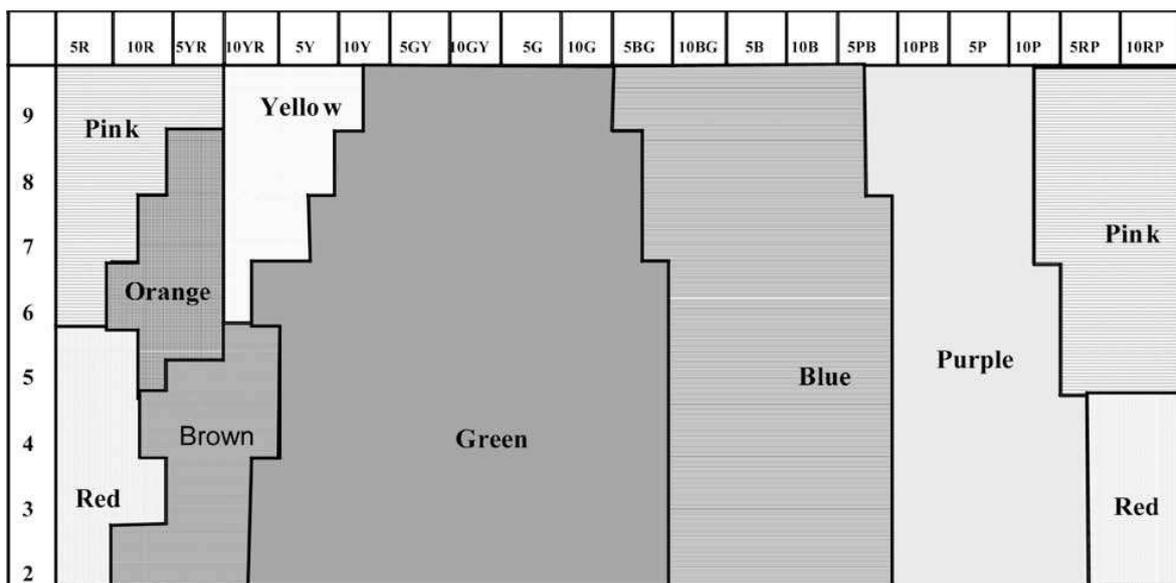
Acknowledgments

This work was supported by European Commission Sixth Framework Program, Grant 012984, Stages in the Evolution and Development of Sign Use (SEDSU). J.G. was supported by an Economic and Social Research Council (ESRC) studentship. We are grateful to Saskia Sabelus and K. Jakurama for helping in the field.

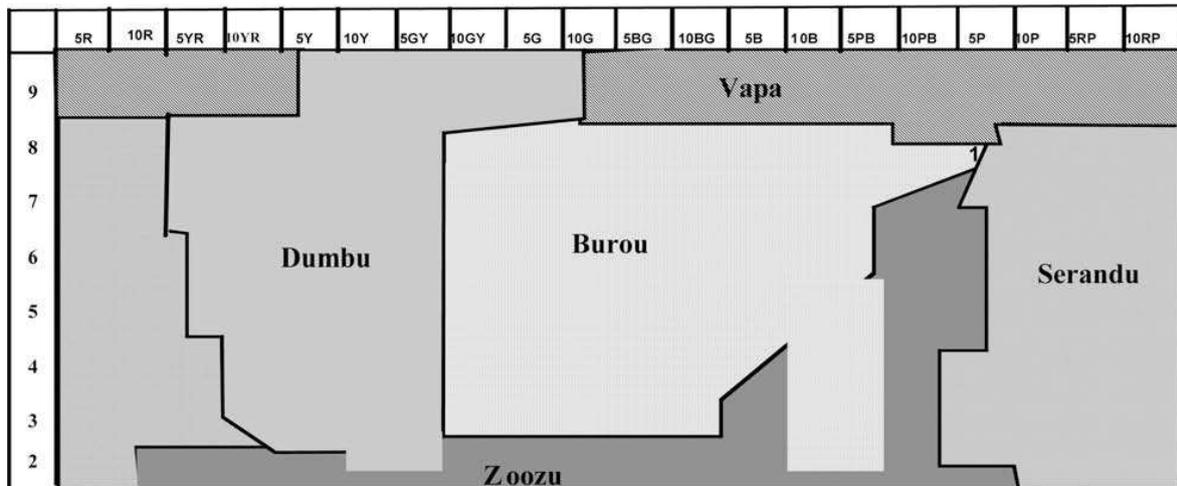
Appendix A

Distribution of English and Himba naming patterns for colors ranging from red–purple on the horizontal axis and from dark (2) to light (9) on the vertical axis.

English color naming distribution



Himba color naming distribution



Note. In the area that English speakers call *purple*, Himba speakers extend their dark term (*zoozu*) up into a range of much paler colors.

Appendix B

Designations of the 22 Color-Aid sweaters in the CIE L*u*v* metric under Illuminant C as measured by Roberson et al. (2005)

Tile code	Color-Aid	Name	L*	u*	v*
1	Gray 4	Gray	71	1	1
4	White	White	100	0	0
6	O S3	Brown	44	13	8
23	YO HUE	Orange	82	101	62
28	V HUE	Purple	40	21	-31
29	BV HUE	Blue	36	-2	-23
33	RVR HUE	Pink–purple	57	113	1
42	Black	Black	35	5	2
45	Y HUE	Yellow	90	54	80
46	YOY HUE	Yellow	92	83	78
47	O HUE	Orange	75	146	44
49	RO HUE	Red	59	136	28
50	B HUE	Blue	49	-30	-69
53	GBG HUE	Green	51	-46	-14
54	G HUE	Green	57	-55	4
59	R T4	Pink	89	70	12
77	ORO T2	Pink–red	70	113	28
78	YGY T1	Green	74	-26	60
79	OYO T4	Orange	83	100	49
80	YOY T3	Orange–yellow	87	78	57
81	RVR T3	Purple	63	81	-13
82	ROR T2	Pink	70	108	15

References

- Berlin, B., & Kay, P. (1969). *Basic color terms: Their universality and evolution*. Berkeley: University of California Press.
- Bornstein, M. (1985). On the development of naming in young children: Data and theory. *Brain and Language*, 26, 72–93.
- Bornstein, M., Kessen, W., & Weiskopf, S. (1976). Color vision and hue categorization in young infants. *Journal of Experimental Psychology: Human Perception and Performance*, 2, 115–129.
- Daoutis, C., Franklin, A., Riddett, A., Clifford, C., & Davies, I. R. L. (2006). Categorical effects in children's color search: A cross-linguistic comparison. *British Journal of Developmental Psychology*, 23, 1–29.
- Davidoff, J. (2001). Language and perceptual categories. *Trends in Cognitive Science*, 5, 382–387.
- Davidoff, J., Davies, I., & Roberson, D. (1999). Color categories in a Stone-Age tribe. *Nature*, 398, 203–204.
- Davies, I. R. L., & Franklin, A. (2002). Categorical similarity may affect color pop-out in infants after all. *British Journal of Developmental Psychology*, 20, 185–203.
- Drivonikou, V., Kay, P., Regier, T., Ivry, R., Gilbert, A., Franklin, A., & Davies, I. R. L. (2007). Further evidence of Whorfian effects to the right visual field. *Proceedings of the National Academy of Sciences of the United States of America*, 104, 1097–1102.
- Fonteneau, E., & Davidoff, J. (2007). Neural correlates of colour categories. *Neuroreport*, 18, 1323–1327.
- Franklin, A., Clifford, A., Williamson, E., & Davies, I. R. L. (2005a). Color term knowledge does not affect categorical perception of color in toddlers. *Journal of Experimental Child Psychology*, 90, 114–141.
- Franklin, A., & Davies, I. R. L. (2004). New evidence for infant color categories. *British Journal of Developmental Psychology*, 22, 349–377.
- Franklin, A., Drivonikou, G. V., Bevis, L., Davies, I. R. L., Kay, P., & Regier, T. (2008). Categorical perception of color is lateralized to the right hemisphere in infants, but to the left hemisphere in adults. *Proceedings of the National Academy of Sciences of the United States of America*, 105, 3221–3225.
- Franklin, A., Pilling, M., & Davies, I. R. L. (2005b). The nature of infant color categorization: Evidence from eye-movements on a target detection task. *Journal of Experimental Child Psychology*, 91, 227–248.
- Gabriel, D. J. (2001). *The spirit of childhood*. Cambridge, UK: RSC Press.
- Gerhardstein, P., Renner, P., & Rovee-Collier, C. (1999). The roles of perceptual and categorical similarity in color pop-out in infants. *British Journal of Developmental Psychology*, 17, 403–420.
- Gilbert, A. L., Regier, T., Kay, P., & Ivry, R. B. (2006). Whorf hypothesis is supported in the right visual field but not the left. *Proceedings of the National Academy of Sciences of the United States of America*, 103, 489–494.
- Glucksberg, S., Hay, A., & Danks, J. H. (1976). Words in utterance contexts: Young children do not confuse the meanings of "same" and "different." *Child Development*, 47, 737–741.
- Goldstein, J., & Davidoff, J. (2008). Categorical perception of animal patterns. *British Journal of Psychology*, 99, 229–243.
- 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, UK: Cambridge University Press.
- Heider, E. R. (1972). Universals in color naming and memory. *Journal of Experimental Psychology*, 93, 10–20.
- Kay, P., & McDaniel, C. K. (1978). The linguistic significance of the meanings of basic color terms. *Language*, 54, 610–646.
- Kay, P., & Regier, T. (2003). Resolving the question of color naming universals. *Proceedings of the National Academy of Sciences of the United States of America*, 100, 9085–9089.
- Kay, P., & Regier, T. (2006). Language, thought, and color: Recent developments. *Trends in Cognitive Sciences*, 10, 51–54.
- MacEvoy, B. (2005). Modern color models: Munsell color system. Color Vision. Available from <http://www.handprint.com/HP/WCL/color7.html#MUNSELL>.
- Munnich, E., & Landau, B. (2003). The effects of spatial language on spatial representation: Setting some boundaries. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and thought*. Cambridge, MA: MIT Press.
- Munsell, A. H. (1905). *A color notation*. Boston: G.H. Ellis.
- O'Hanlon, C. G., & Roberson, D. (2006). Learning in context: Linguistic and attentional constraints on children's color term learning. *Journal of Experimental Child Psychology*, 94, 275–300.
- Pitchford, N. J., & Mullen, K. (2001). Conceptualization of perceptual attributes: A special case for color? *Journal of Experimental Child Psychology*, 80, 289–314.
- Pitchford, N. J., & Mullen, K. (2002). Is the acquisition of basic color terms in young children constrained? *Perception*, 31, 1349–1370.
- Pitchford, N. J., & Mullen, K. T. (2003). The development of conceptual color categories in preschool children: Influence of perceptual categorization. *Visual Cognition*, 10, 51–77.
- Pitchford, N. J., & Mullen, K. T. (2005). The role of perception, language, and preference in the developmental acquisition of basic color terms. *Journal of Experimental Child Psychology*, 90, 275–302.
- Regier, T., Kay, P., & Cook, R. S. (2005). Focal colors are universal after all. *Proceedings of the National Academy of Sciences of the United States of America*, 102, 8386–8391.
- 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., 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, 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., Davies, I., & Davidoff, J. (2000). Color categories are not universal: Replications and new evidence from a Stone-Age culture. *Journal of Experimental Psychology*, 129, 369–398.
- Roberson, D., Pak, H. S., & Hanley, J. R. (2008). Categorical perception of color in the left and right visual field is verbally mediated: Evidence from Korean. *Cognition*, 107, 752–762.
- Rosch, E. (1973). Natural categories. *Cognitive Psychology*, 4, 328–350.

- Soja, N. (1994). Young children's concept of color and its relation to the acquisition of color words. *Child Development*, 65, 918–937.
- Tan, L. H., Chan, A. H. D., Kay, P., Khong, P-L., Yip, L. K. C., & Luke, K.-K. (2008). Language affects patterns of brain activation associated with perceptual decision. *Proceedings of the National Academy of Sciences of the United States of America*, 105, 4004–4009.
- Winawer, J., Witthoft, N., Frank, M. C., Wu, L., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences of the United States of America*, 104, 7780–7785.