



## New evidence for infant colour categories

Anna Franklin and Ian R. L. Davies\*

University of Surrey, UK

Bornstein, Kessen, and Weiskopf (1976) reported that pre-linguistic infants perceive colour categorically for primary boundaries: Following habituation, dishabituation only occurred if the test stimulus was from a different adult category to the original. Here, we replicated this important study and extended it to include secondary boundaries, with a crucial modification: The separations between habituated and novel stimuli were equated in a perceptually uniform metric (Munsell), rather than in wavelength. Experiment 1 found Categorical Perception and no within-category novelty preference for primary boundary blue–green and secondary boundary blue–purple. Experiment 2 replicated the categorical effect for blue–purple and found no within-category novelty preference with increased stimulus separation. Experiment 3 showed category effects for a lightness/saturation boundary, pink–red. Novelty preference requires a categorical difference between the habituated and novel stimulus. The implications for the origin of linguistic colour categories are discussed.

Colour perception is categorical. Although the colour spectrum is continuous, it appears to be segmented into qualitatively different perceptual categories. Moreover, discrimination of pairs of colours from adjacent categories is easier than discrimination of equivalently spaced colours in the same category. This is the hallmark of ‘Categorical Perception’ (Harnad, 1987). In English, these perceptual categories are also distinguished linguistically by category names such as *red*, *green*, *yellow*, and *blue*. The relationship between these perceptual and linguistic categories is in dispute. For some (Universalists), perceptual categories are ‘hard-wired’ into the visual system, and language exploits these discontinuities in perceptual colour space. For others (Relativists), perceptual categories are constructed through language, and colour perception varies as language varies.

The clearest evidence for the hardwiring of perceptual categories comes from a classic study by Bornstein, Kessen and Weisskopf (1976). Following habituation to say, red1, 4-month-old infants looked longer at a colour from a different adult category (yellow) than one from the same category (red2), even though the differences in wavelength of the test stimuli from the standard were equal. Their results suggest that

\* Correspondence should be addressed to Professor Ian Davies, Department of Psychology, University of Surrey, Guildford GU2 7XH, UK (e-mail: [i.davies@surrey.ac.uk](mailto:i.davies@surrey.ac.uk)).

pre-linguistic infants have perceptual colour categories, and show Categorical Perception by at least 4 months of age. Despite the importance of this result, and contemporaneous criticism of the methods (e.g. Werner & Wooten, 1985), it has never been replicated. The first aim of the present set of studies was to do so.

The corner stone of the relativist's case is that linguistic colour categories vary across languages (Berlin & Kay, 1969). If linguistic categories are determined by universal perceptual categories, why should they vary so? One possibility, that we investigate here, is that primary category boundaries (red, green, yellow and blue) are hardwired, but secondary categories (brown, orange, pink and purple) are not. Experiment 3 also tested whether infants show Categorical Perception across a boundary defined by lightness and saturation, rather than hue, and explored the relationship among hue, saturation and lightness in infants.

### **Language variations**

The most extensive and systematic study of languages' colour terms has been conducted by Berlin and Kay and coworkers (Berlin & Kay, 1969; Kay, Berlin, & Mayfield, 1991; Kay & Maffi, 1999; Kay & McDaniel, 1978). Berlin and Kay (1969) argued that, although languages varied in their numbers of colour terms, they did so in a systematic way. These regularities were encapsulated in the following hierarchy (terms bracketed together share common positions)

[BLACK & WHITE] RED [GREEN OR YELLOW] BLUE BROWN [PURPLE PINK ORANGE GREY]

The lower the position on the hierarchy (further to the right) the smaller the number of languages that encode a term. For instance, most languages encode BLACK, WHITE and RED, whereas relatively few encode PURPLE, PINK, ORANGE and GREY.

The first six terms (primaries) correspond with Hering's (1878) unique hues and to the cardinal directions in colour space (e.g. Krauskopf, Williams, & Heeley, 1982). The remaining terms (secondaries) appear phenomenally to be blends of pairs of primaries. Despite early optimism (see De Valois & Jacobs, 1968; Kay & McDaniel, 1978), the neurological basis for the primary-secondary distinction remains to be uncovered (De Valois & De Valois, 1990). Nevertheless, languages are more likely to encode primary categories than secondary categories. This leads to the conjecture that we test here, that primary categories are hardwired, whereas secondary categories are learned. However, as the hierarchy implies, primary categories are also ordered: for instance, BLUE is less common than GREEN, YELLOW OR RED. Thus, if primary categories are hardwired, but secondary categories are not, this will not explain all of the linguistic variation.

### **Categorical Perception**

Categorical Perception is shown by better between-category than within-category discrimination. This holds even when between and within-category pairs are equated in their separations in a perceptually uniform metric. For instance, the Munsell colour order system is standardized so that stimuli separated by the same number of hue steps have the same perceptual distance throughout the colour space. Bornstein and Korda (1984), using a successive same-different task, found that discrimination across the blue-green boundary was faster than intra-green and intra-blue pairs, even though all pairs were the same number of hue steps apart. This finding is robust and is found with different boundaries (e.g. Pilling & Davies, in press), different colour-order systems (e.g.

Boynton, Fargo, Olson, & Smallman, 1989) and different methods (e.g. Roberson & Davidoff, 2000).

Categorical Perception does not mean that category exemplars are indistinguishable from each other. Rather, discriminability varies with position in the category, peaking across category boundaries. It is as though perceptual colour space has been transformed topologically or 'warped' (Harnad, 1987). The transformation stretches perceptual distances across category boundaries relative to within-category distances (e.g. Goldstone, 1994; Harnad, 1987). Discrimination should become easier as perceptual distance increases, and if the separation includes a category boundary.

There are a number of converging reasons to suppose that Categorical Perception of colour is partly learned. First, sensitivity to many stimulus dimensions can be improved by training (see Fahle & Poggio, 2002), and this includes colour discrimination (Özgen & Davies, 2002). Such learning could induce Categorical Perception by discrimination training around the category boundary. Second, consistent with this, Özgen and Davies (2002) found that learning new categories, such as dividing green into two, induced Categorical Perception across the new boundary. Third, Categorical Perception varies with the positions of languages' category boundaries. For instance, Categorical Perception is not found across the blue-green boundary in speakers of languages that do not mark that distinction (Kay & Kempton, 1984; Roberson, Davies & Davidoff 2000). Taken together, these three strands of evidence are consistent with the possibility that the location and strength of Categorical Perception may be determined, at least partly, during language learning. These processes may modify hardwired structures or act on a tabula rasa.

### **Infant colour vision and categorization**

The signature of colour vision is discrimination of stimuli differing in hue at constant luminance (lightness or brightness) and saturation. Infants have some chromatic discrimination by at least 3-months, with the sensitivity to blue being the slowest to develop (Adams, Courage, & Mercer, 1994; Teller, 1998). However, chromatic thresholds are considerably higher than for adults and do not reach adult levels until the teenage years (Teller, 1998). In the metrics used in the current paper, infant thresholds are of the order of 10  $\Delta E$  units (CIE 1976) or about one Munsell hue step.<sup>1</sup> Despite the overall difference in chromatic sensitivity, recent studies suggest that the shape of the infant spectral sensitivity function is similar to that for adults. This means that if stimuli are at adult isoluminance, they will probably also be isoluminant for infants (Knoblauch, Vital-Durand, & Barbur, 2001; Pereverzeva, Chien, Palmer, & Teller, 2002). This implies that Munsell stimuli equated in value (lightness) can be used in the study of infant colour vision with reasonable safety.

The strongest evidence for infant Categorical Perception of colour comes from Bornstein *et al.*'s (1976) classic study. Using monochromatic lights, 4-month old infants were habituated to stimuli of one wavelength, and dishabituation rates to the original stimulus and novel stimuli were compared. The novel stimuli were either from a different adult category to the original or from the same category, and the stimuli were equated in luminance (3.4 cd/m<sup>2</sup>). The blue-green, green-yellow and yellow-red boundaries were investigated. For each category boundary, the within and between

<sup>1</sup>  $\Delta E$  is the Euclidean distance between two point in the CIE colour space  $Lu^*v^*$ .  $L$  is lightness,  $u^*$  is the red-green axis and  $v^*$  the blue-yellow axis (see Hunt, 1987).

pairs were separated by equal differences in wavelength, but the size of these separations varied with category boundary. These separations were 30, 10, and 20 nm for the blue–green, green–yellow and yellow–red boundaries, respectively. In general, looking times for new category stimuli were longer than for same category stimuli, and for the original stimulus, whereas looking times for the habituated stimulus and novel same category stimuli did not differ. Infants responded differentially to a spectral change that included a category shift, but not to a spectral change alone: they treated within-category stimuli as similar and between-category stimuli as different. There were exceptions to this (one of the red–yellow sets), nevertheless, Bornstein *et al.* concluded that infants perceive the colour continuum categorically. Bornstein *et al.* also looked for a possible categorical difference between blue and violet—a secondary boundary in our terms—between 450 and 420 nm. Infants behaved as though blue and violet were in the same category. However, the basic secondary category is purple rather than violet, and purple is not a spectral colour. Rather, it consists of a mixture of short and long wavelengths (blue and red). It is possible, therefore, that infants do have a blue–purple (and other) secondary category boundary, and we explore that possibility here.

### **Limitations of Bornstein *et al.*'s study**

Werner and Wooten (1985) argued that the infant discriminations could have been based on residual differences in brightness and saturation rather than on hue. The stimuli were at adult isoluminance, and, as mentioned earlier, recent evidence (Pereverzeva *et al.*, 2002) suggests that they would have also been isoluminant for infants. However, the issue of residual saturation differences remains. Monochromatic lights, as used by Bornstein *et al.*, do vary in their saturation levels, and this could have influenced the results to some degree. Moreover, such stimuli are the most saturated possible instances of a given hue, and these levels do not normally occur in the natural world (nor in the world in which colour vision evolved). Thus, it is possible that infants show Categorical Perception at these extreme levels of saturation but may not at more typical levels. Similarly, most natural stimuli are reflective, rather than radiant, and consist of broad band mixtures of wavelengths, rather than the monochromatic stimuli used by Bornstein *et al.* Given the significance of the results, it is important to test whether they extend to stimuli more typical of the natural world.

Finally, and most crucially, there is an issue over what the appropriate metric for equating within and between-category separations should be. Bornstein *et al.*'s separations were equal in wavelength for a given boundary. However, wavelength discrimination thresholds vary with position in the spectrum (e.g. see Wyszecki & Stiles, 2000, p. 571). This means that equal wavelength separations are not necessarily at equal perceptual separations, as measured, for instance, by number of JNDs (Wyszecki & Stiles, 2000, p. 572). If looking times are determined by the perceptual difference between the habituated stimulus and the test stimulus, then the inequality in perceptual distance might account for the results.<sup>2</sup> However, we show below that stimulus separation alone, whether measured in wavelength or a measure of perceptual distance (CIE  $\Delta E$ ) does not predict looking time.

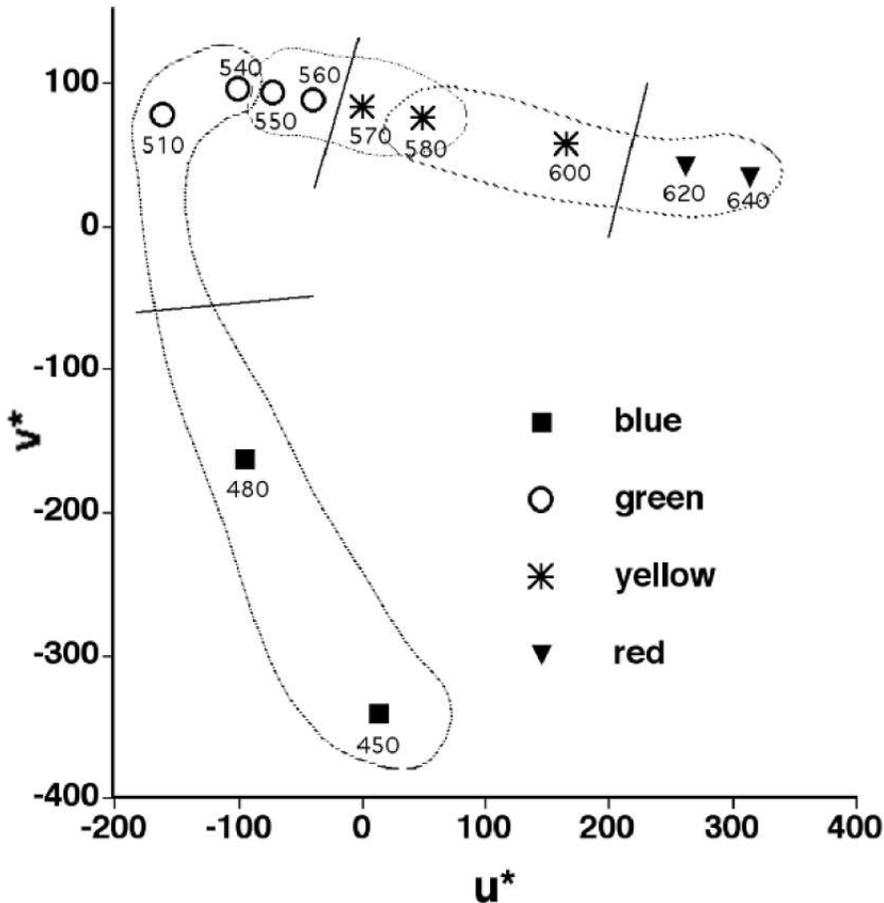
<sup>2</sup> If this were the origin of the effect, it would not really be equivalent to adult categorical perception. Equivalent non-uniformities also occur within categories. Thus, quasi-categorical effects should also be found within categories. It might be thought that it is this variation in discriminability that drives category formation. Category boundaries occur at local discrimination maxima. This fits the blue–green boundary; there is a discrimination peak at about the right place. But, it is much less clear for green–yellow and yellow–orange–red.

**Table 1.** Wavelength ( $\lambda$ ) and adult category of stimuli used by Bornstein *et al.*, the separation of stimulus pairs in  $\lambda$  and CIE ( $\Delta E$ ), and associated looking times.

Habituated stimulus		Test stimulus		Distance		Looking time (s)
$\lambda$ (nm)	Category	$\lambda$ (nm)	Category	$\lambda$	(CIE)	
480	Blue	480	Blue	0	0.0	5.7
		450	Blue	30	203.5	5.8
		510	Green	30	243.3	7.3
510	Green	510	Green	0	0.0	5.0
		540	Green	30	252.5	4.0
		480	Blue	30	243.3	7.0
560	Green	560	Green	0	0.0	3.8
		550	Green	10	32.7	3.8
		570	Yellow	10	40.1	6.4
570	Yellow	570	Yellow	0	0.0	4.3
		580	Yellow	10	48.6	3.3
		560	Green	10	40.1	6.2
600	Yellow	600	Yellow	0	0.0	4.4
		580	Yellow	20	114.5	4.1
		620	Red	20	94.7	6.0
620	Red	620	Red	0	0.0	5.4
		640	Red	20	50.9	6.0
		600	Yellow	20	94.7	4.4

Table 1 shows Bornstein *et al.*'s looking times taken from the text of the paper or from the graph on p. 119. It also gives the wavelength of each stimulus, and the stimulus separations in wavelength and  $\Delta E$ . For each habituated stimulus, there are three test stimuli: the repeated habituated stimulus; a novel within-category stimulus; and a novel between-category stimulus. Consider the wavelength metric first. Within-category novel stimuli are looked at no longer than the habituated stimulus. That is, a separation well above threshold (10, 20 or 30 nm) engenders no more looking than a zero separation. Moreover, it is apparent from comparing the three colour regions tested that the size of the separation makes no difference. Looking times for novel within-category stimuli are shorter than for novel between-category stimuli, but the extent of this does not vary with varying stimulus separations (10, 20 or 30 nm). Similarly, there appears to be no systematic association between looking times for novel between-category stimuli and the size of the separation (again, 10, 20 or 30 nm). The infants behave as though they require a category shift to engage their interest, rather than novelty alone.

Transforming the colour space into a perceptually uniform metric does not explain the lack of effect of separation size. Figure 1 shows the location of the stimuli in CIE (1976) colour space ( $u^*$ ,  $v^*$ ). The separations in the diagram represent perceptual distances ( $\Delta E$ ; recall that luminance did not vary, and thus there are no differences in  $L$ ), and Table 1 gives the separations for each pair in  $\Delta E$  alongside the wavelength separation. It can be seen that there are still differences in the separation sizes for the three regions in  $\Delta E$ . However, now, within regions, the separations are also not equal. For instance, for the blue–green region, the between-category distance (243.3) is larger than one within-category separation (203.5) and smaller than the other within-category



**Figure 1.** Location of Bornstein *et al.*'s stimuli in the CIE colour space ( $u^*$ ,  $v^*$ ), with stimulus wavelength given. Symbols indicate category, straight lines indicate category boundaries, and dotted lines indicate stimulus sets.

separation (252.2), and the latter two are not equal. For all the regions, one within-category distance is larger than the between-category distance, and one smaller. Again, these inequalities do not appear to affect looking times.

The apparent lack of a perceptual distance effect and the indifference to novel same-category stimuli are puzzling. They are inconsistent with the perceptual warping account of Categorical Perception outlined earlier. Rather than perceptual distance and the category effect being additive, infant looking seems to be determined by the categorical difference alone. In the study reported here, we manipulate perceptual distance to check the generality of these results.

### Recent studies

Gerhardstein, Renner and Rovee-Collier (1999) investigated effects of categorical and perceptual similarity in infant colour 'pop-out'. The logic and stimulus structure was

similar to that of Bornstein *et al.*, but they avoided some of the problems inherent in using monochromatic stimuli just described. They used reflective (Munsell) stimuli at constant saturation (chroma), and the stimulus separations were equated in Munsell hue steps—equal perceptual distances rather than equal wavelengths. They tested a secondary boundary (red–orange), and the discrimination task differed from that used by Bornstein *et al.*'s. They familiarized infants to the standard colour (e.g. red1) and then tested whether a new colour was detectable (popped out) among six instances of the standard colour. The new colour was either in the same adult category as the standard (red2) or from the adjacent category (orange), and the perceptual distance of the new stimuli from the standard also varied (either two or three Munsell hue steps). They report that pop-out increased with perceptual distance, but the categorical relationship had no effect. Note that this is the opposite pattern to that found by Bornstein *et al.* This could be for a number of reasons. First, different processes might be involved when viewing multiple stimuli (Gerhardstein *et al.*) and single stimuli (Bornstein *et al.*). For instance, perhaps pop-out is driven by perceptual distance or the 'physical code', whereas habituation may be driven primarily by the 'categorical code' (Bornstein & Korda, 1984). Second, it could be that Bornstein *et al.*'s results were influenced by saturation differences, and the category effect does not occur at constant saturation (chroma). Third, infants, unlike adults, may not show Categorical Perception when the separations are in a perceptually uniform metric (Munsell). Fourth, it could be that infants do not perceive reflective (broad band) or low saturation stimuli categorically. Fifth, perhaps infants may not have an orange category; Bornstein *et al.* tested the red–yellow boundary, not the red–orange, and even so, the signs of Categorical Perception were the weakest of the three boundaries they tested. However, there were also methodological problems with Gerhardstein *et al.*'s study. Incandescent light was used rather than the illuminant under which Munsell colours are standardized. This use of incandescent light produced unequal perceptual distances, undermining their experimental design. There were also problems with the categorical status of the stimuli. At least one intended within-category (red) pair was between-category (pink–red). Therefore, the question of whether infants show Categorical Perception with reflective stimuli in perceptually uniform metrics, as adults do, is unresolved (see Davies & Franklin, 2002),<sup>3</sup> and Bornstein *et al.*'s study remains the only evidence for Categorical Perception of colour in infants.

### **The current studies**

The main aim of the current studies was to see if infants showed signs of Categorical Perception at 4 months when the stimuli were spaced in a perceptually uniform metric. We also tested the conjecture that this would be the case for primary categories but not for secondary categories. Experiment 1 tested the blue–green and blue–purple boundaries and also varied the size of the stimulus separations. Infants showed signs of Categorical Perception for both boundaries, but there were no effects of the size of the separations. Experiment 2 retested the blue–purple boundary and increased the

<sup>3</sup> An earlier study by Fagan (1974) was similarly undermined by using the wrong illuminant (see Wooten, 1974). Although not specifically about infant categorical perception, Fagan's data suggest that infants were indifferent to checkerboards made up of two Munsell colours, if both colours were from the same adult category, whereas they looked preferentially at them if they were from different categories. However, it appears that use of incandescent light reduced the difference among the within-category stimuli to below infant threshold.

separation further. Experiment 3 extended the investigations to include another secondary boundary (pink-red) where the boundary was defined by lightness and saturation, rather than hue.

## EXPERIMENT I

### Categorical Perception for blue-green and blue-purple

Infants were familiarized with stimuli of one hue and then tested for novelty preference with the original hue and a novel hue simultaneously. Stimulus pairs varied in their categorical relationship (between/within-category), as verified by adult judgments, and their perceptual distance (near/far). Greater novelty preference for novel new category stimuli than for novel same category stimuli was taken as indicating Categorical Perception. Greater novelty preference for far pairs than near pairs was taken as indicating a perceptual distance effect. The stimuli were reflective Munsell stimuli, varied only in hue and were of a lower saturation than Bornstein *et al.*'s lights. This was to check that Bornstein *et al.*'s results generalized to these more natural stimuli and that the original results were not due to saturation differences. Stimulus separations were equated in the Munsell perceptually uniform metric with appropriate illumination (CIE illuminant C) to see if the lack of categorical effects in Gerhardstein *et al.*'s data was due to the lighting. Finally, the size of the stimulus separations was varied to see if perceptual distance effects and categorical effects were additive, as the perceptual warping hypothesis suggests.

## Method

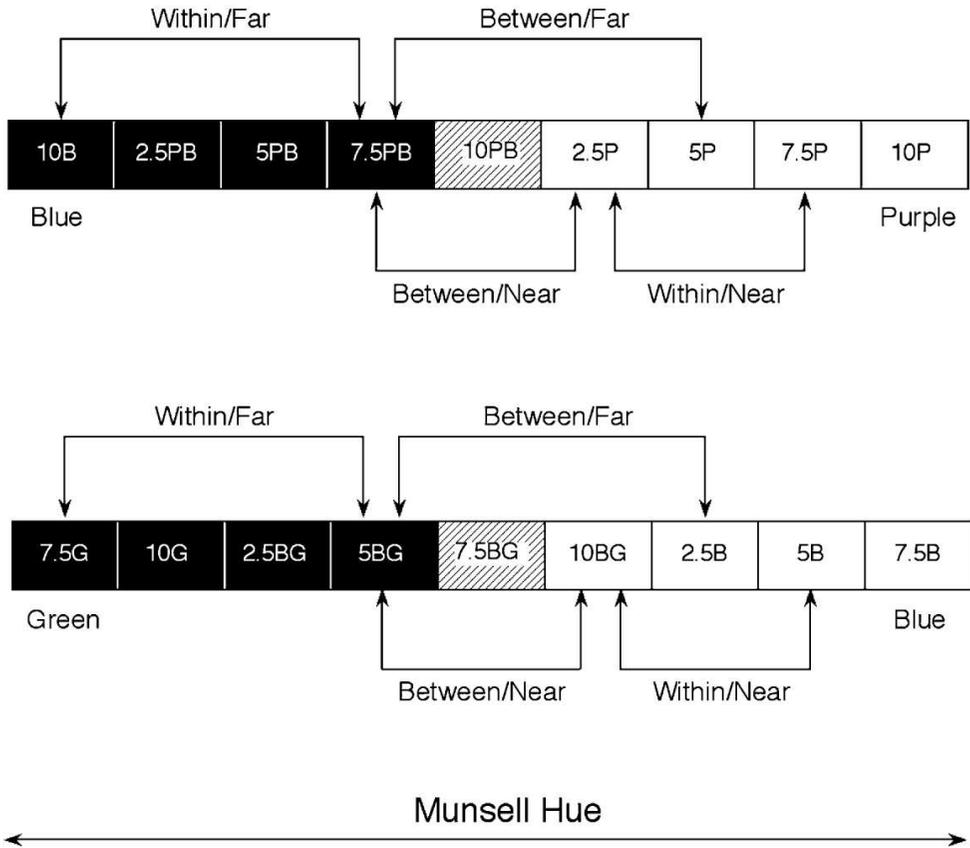
### Participants

Fifty-five infants were recruited from baby groups, nurseries and the local area through posters, flyers and word of mouth. Nineteen infants were not included in the final study, 15 due to general fussiness, 1 due to falling asleep during testing, 2 due to illness and 1 due to their parent reporting a family history of colour blindness. The mean age of the 36 infants in the final sample was 17 weeks (range: 16–20 weeks). Twenty-one infants in the final sample were female, and 15 were male.

Ten adults from the University of Surrey (five men and five women) took part in the naming and similarity judgments task. All adults had normal colour vision, as indicated by the City University Test (Fletcher, 1981).

### Materials and stimuli

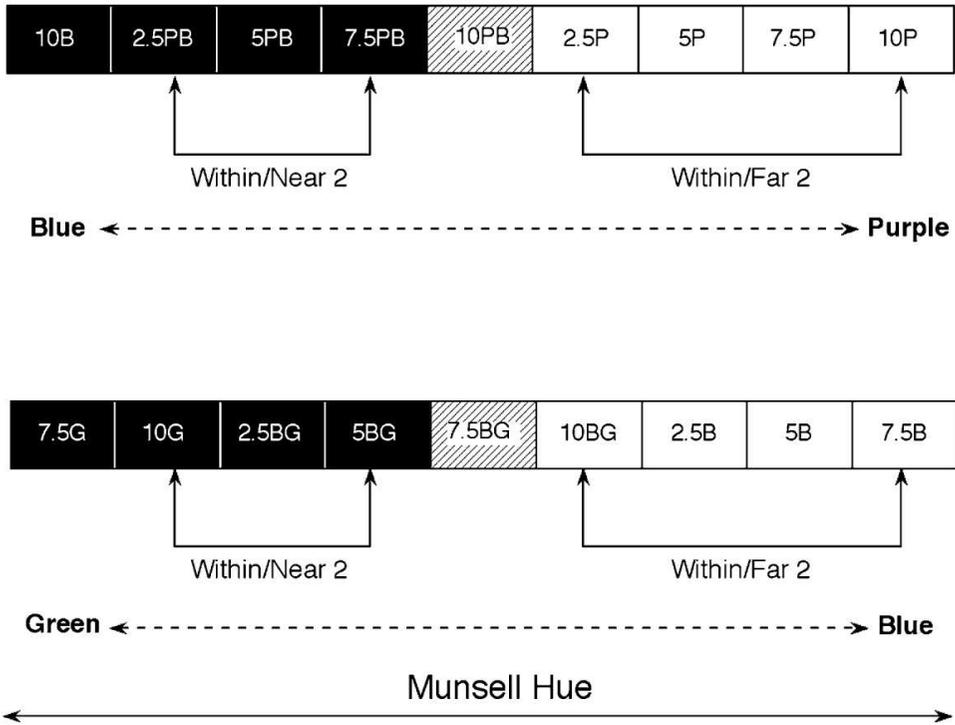
Stimuli were glossy squares of Munsell paper mounted on card from the blue-purple and the blue-green regions. There were four main pairs of stimuli varying in categorical status (within or between-category) and in perceptual distance (near or far). Near pairs were separated by two steps of Munsell hue, far pairs by three steps both at constant value and chroma. The hue notations and their relationship are shown in Fig. 2. Two additional within pairs were included to ensure that any within-category effects were not colour-specific (Fig. 3).



**Figure 2.** Munsell codes, categorical status and Munsell distances of the stimuli of the four main experimental pairs used in Experiment 1. The relative perceptual distances (near/far) and categorical relationships (within/between) of the experimental pairs are also shown. The shaded cell indicates an adult category boundary.

**Display**

The display is illustrated in Fig. 4. Infants were seated in a baby car-seat 40 cm away from a black screen with two 12 cm<sup>2</sup> (17°) stimulus ‘windows’, 12 cm (17°) apart edge to edge. A video camera recorded the infant’s face through a small window obscured by black gauze, between the two stimulus windows. A small light embedded in the screen above the camera window was turned on at the onset of a trial to encourage central fixation by the infant. This could be assessed by inspecting the video output on a monitor behind the screen. Between trials, the screen was covered by a black curtain. Stimuli were illuminated by two projectors placed 95 cm behind the car seat at a 45° angle to each stimulus window. These provided the only source of illumination and were fitted with Lee Filters to produce illuminant C (6500K at 1100 lux).



**Figure 3.** Munsell codes, categorical status and Munsell distances of the stimuli of the additional within pairs used in Experiment 1. The relative perceptual distances (near/far) and categorical relationships (within/between) of the experimental pairs are also shown.

**Procedure**

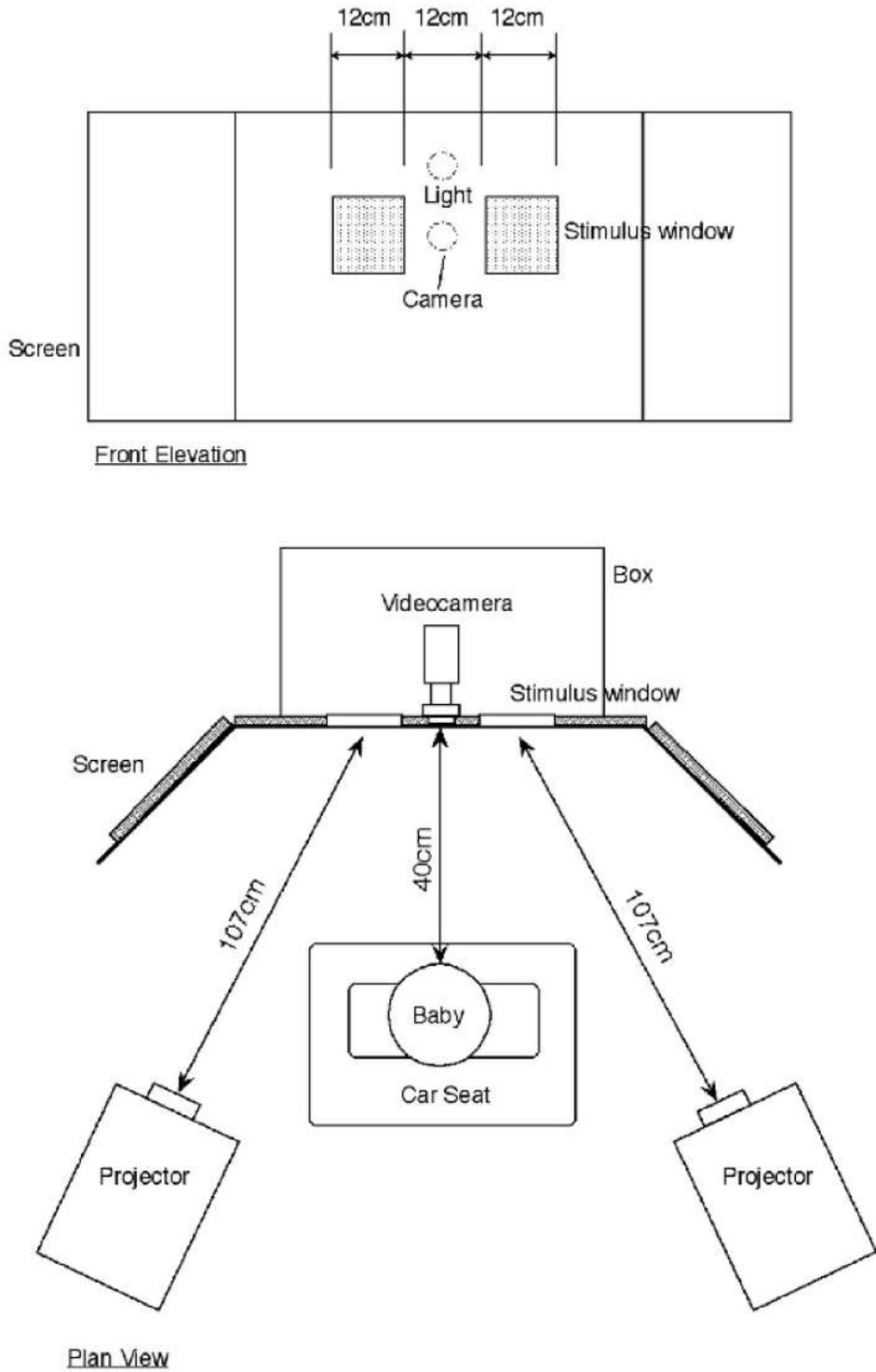
*Naming and similarity judgments*

In the naming task, adults were presented with stimuli individually on a grey cloth under illuminant C and were asked to name the colour. In the similarity-judgement task, adults were shown each experimental pair of stimuli and asked to rate their similarity/dissimilarity on an analogue scale. The far right of the line indicated a high dissimilarity, and the far left indicated a low similarity. The order of tasks and the order of stimulus presentation were randomized.

*Infant procedure*

There were six stimulus pairs for each colour region (Figs 2 and 3). Six infants, half boys and half girls, were allocated randomly to each pair, with the constraint that the groups were matched as closely as possible for age. Each infant was tested on one stimulus pair for each colour region. Thus, all key comparisons (between/within-category; near/far) involved separate groups.

Infants were tested using a paired-comparison, novelty preference method (see Catherwood, Crassini & Freiberg, 1990). For each stimulus pair, one stimulus was the familiarization stimulus, and the other was the novel stimulus. The allocation of familiar and novel stimulus was split so that for half of the infants in a condition, one stimulus in



**Figure 4.** Scaled plan and elevation of the experimental set-up.

the pair was the familiarization stimulus, and for the other half, it was novel. During the familiarization phase, one stimulus from a pair (familiarized stimulus) was presented in both stimulus windows for ten 8-s trials with 5-s between trials. After a 5-s delay, the test phase followed in which the remaining stimulus from the pair (novel stimulus) was presented alongside the familiarized stimulus for five 8-s trials with 5-s intervals. The start and end of the trials were indicated by a 'beep'. The left/right position of the novel stimulus during the test phase was randomized across trials. At the end of the session, a small rattling toy was waved in front of each stimulus window to encourage fixation, which was recorded and used as a template when scoring the video images for each trial.

## Results

### Adult naming and similarity judgments

The pattern of naming was 90–100% consistent with the intended nominal categories for both sets. Table 2 shows across participant mean dissimilarity judgments (measured in millimetres) for each combination of categorical status and perceptual distance, for blue–green and blue–purple separately. It appears that between-category pairs were judged as more dissimilar than within-category pairs and that far pairs were judged as more dissimilar than near pairs.

**Table 2.** Means (SD) of adult dissimilarity judgments (mm) for the four experimental pairs for each colour region

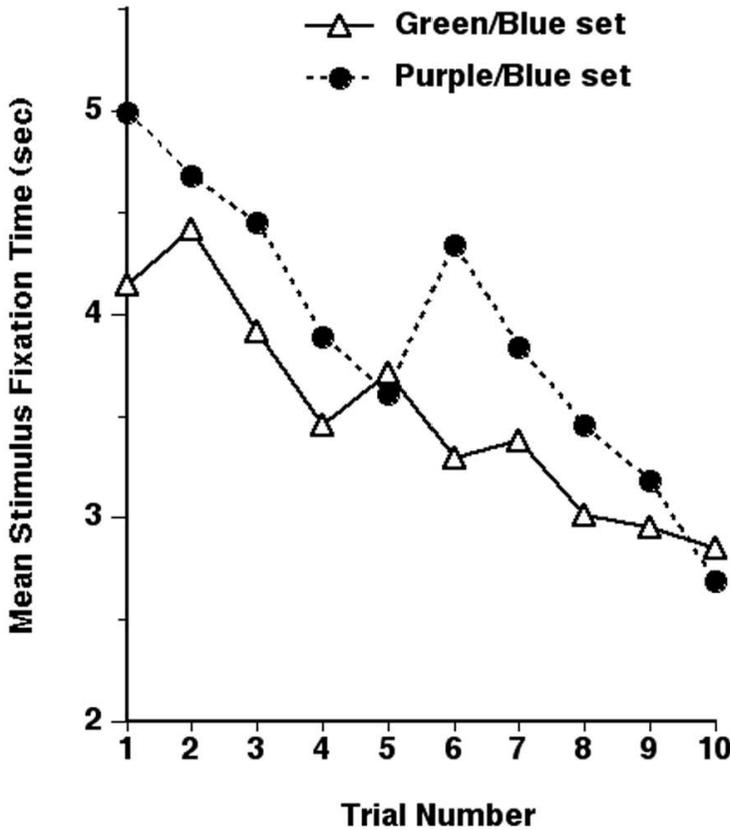
Category	Blue–green		Blue–purple	
	Near	Far	Near	Far
Within	30.90 (17.92)	36.40 (19.55)	36.70 (24.72)	51.40 (28.40)
Between	43.50 (19.71)	59.50 (16.89)	67.70 (18.40)	73.80 (14.05)

This was supported by two repeated-measures ANOVAs on dissimilarity ratings (mm) with category (between/within) and distance (near/far) as factors. There were significant main effects of category and distance for both sets: category blue–green,  $F(1, 9) = 13.77, p < .005$ ; category blue–purple,  $F(1, 9) = 14.85, p < .005$ ; distance blue–green  $F(1, 9) = 5.59, p < .05$ ; distance blue–purple,  $F(1, 9) = 8.271, p < .05$ . No interactions were significant (largest  $F = 3.30$ , smallest  $p = 0.10$ ).

### Infant data

Fixations were coded blind to the experimental condition. A second observer coded a subset of data from six randomly selected coding sessions. Pearson product-moment correlations, computed for fixation across the 15 trials of the session, yielded a mean inter-observer reliability of  $r(15) = .89, p < .001$ .

To check for any a priori preferences for hue, separate repeated-measures ANOVAs for each colour set were conducted on the fixation times during the familiarization phase with stimulus (5) and trial (10) as factors. There were no significant effects; largest  $F(4, 19) < 1, p = .86$ .



**Figure 5.** Mean stimulus fixation times during the familiarization phase for blue–green and blue–purple colour sets.

Figure 5 shows the mean stimulus fixation times over the 10 trials of the familiarization phase, for both stimulus sets. Looking times tended to decrease across trials from 4–5-s to less than 3-s. This impression was tested formally with repeated-measures ANOVAs on trials (10) for each colour region. There were significant decreasing linear trends for both regions: blue–green,  $F(1, 23) = 14.28, p < .005$ ; blue–purple,  $F(1, 23) = 23.00, p < .001$ .

Raw test fixation times were weighted by side preference. For each infant, a preference ratio was calculated. This was the ratio of total fixation time on each side during the familiarization phase, with the larger score divided by the smaller score. The test fixation scores on each infant’s preferred side were then divided by this ratio, and these weighted scores were used to calculate novelty preference.

Novelty preference was expressed as the percentage of total fixation time spent fixating the novel stimulus:

$$\text{Novelty preference} = (\text{novel fixation} / (\text{novel fixation} + \text{familiar fixation})) \times 100.$$

A score above 50% indicates a novelty preference. Figures 6A and 6B show the novelty preference scores for the four stimulus pair types, for the blue–green set and for the blue–purple set, respectively. It appears that the scores are all larger than 50% for the four between-category conditions, and around 50% for the six within-category conditions. One-sample two-tailed  $t$  tests support this impression: blue–green: between/near  $t(5) = 3.39$ ,  $p < .025$ ; between/far  $t(5) = 5.61$ ,  $p < .005$ ; blue–purple: between/near  $t(5) = 2.801$ ,  $p < .05$ ; between/far  $t(5) = 3.70$ ,  $p < .025$ . All within-category scores were less than 50% but none significantly so; maximum  $t(5) = -0.64$ ,  $p = .55$ .

A two-way ANOVA on the novelty preference scores<sup>4</sup> with perceptual distance (near/far) and categorical status (between/ within) as factors revealed a significant main effect of category for the blue–green set;  $F(1, 20) = 25.91$ ,  $p < .0001$ , and for the blue–purple set;  $F(1, 20) = 10.95$ ,  $p < .005$ . Main effects of distance and interactions of category and distance were not significant: largest  $F(1, 20) < 1$ ,  $p = .51$ . There were no differences (main effects or interactions) involving colour set: largest  $F(1, 40) = 1.20$ ,  $p = .28$ . Replacing the original within pair scores with the extra within pair scores yields the same pattern: blue–green  $F(1, 20) = 22.72$ ,  $p < .001$ ; blue–purple  $F(1, 20) = 11.12$ ,  $p < .005$ , and the main effect of distance and the interaction between-category and distance remain non-significant.<sup>5</sup>

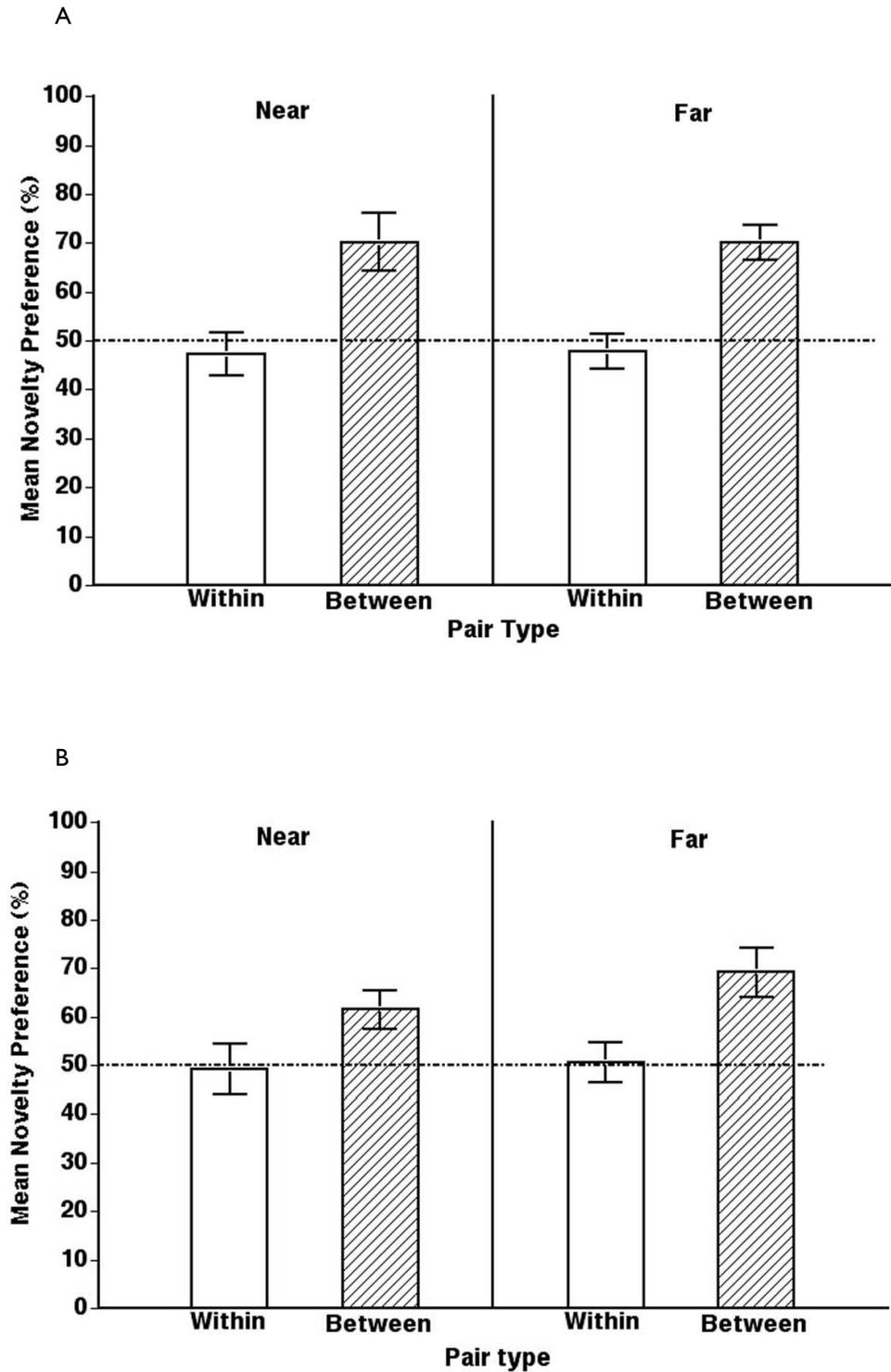
## Discussion

In the preliminary study, naming by adults was consistent with the intended categorical status of the stimuli. Their similarity judgments also showed that far pairs were judged to be more dissimilar than near pairs, consistent with their Munsell separation. Moreover, between-category pairs were judged to be more dissimilar than within-category pairs, consistent with Categorical Perception.

In the main study, infants familiarized with stimuli from the blue–green and blue–purple sets as shown by decreasing fixation time across trials. During the subsequent test phase, they showed no novelty preference for new stimuli from the same adult category (novelty preference scores not different to 50%), whereas it was shown for all between-category pairs (scores all greater than 50%). The size of the novelty preference effect did not differ between the two stimulus sets (blue–green and blue–purple). Nor did the perceptual distance among pairs (near or far) have any effect. There were no a priori preferences that could account for this pattern of results. These data meet our operational definition of Categorical Perception given in the introduction, and imply that pre-linguistic infants can detect the categorical difference among colour stimuli by 4 months of age. The results are consistent with Bornstein *et al.*'s result and show that infant Categorical Perception occurs for reflective stimuli of a lower saturation than Bornstein *et al.*'s, and for stimulus separations equated in a perceptually uniform metric.

<sup>4</sup> The weighted raw scores were analysed using three-way ANOVAs on looking times for familiar-novel, categorical status and perceptual distance as factors. Crucially, the interaction between of familiar-novel and categorical status was significant; green–blue,  $F(1, 20) = 11.76$ ,  $p < .05$ ; purple–blue,  $F(1, 20) = 8.463$ ,  $p < .01$ , indicating that infants looked more at novel colours from a different adult category than at novel colours from the same adult category.

<sup>5</sup> Novelty preference scores were also analysed in terms of age in weeks (16/17/18/19/20) and gender. No significant main effects or interactions were found.



**Figure 6.** Novelty preference for the four experimental pairs in Experiment 1: category (within/ between) and distance (near/far) for the (A) blue–green set and the (B) blue–purple set.

However, these data offer no support to our conjecture that primary categories are hardwired, whereas secondary categories are learned. Both boundaries appear to elicit Categorical Perception, and the strength of these effects did not differ between boundaries. The apparent equivalence of the two boundaries must be treated with caution for two reasons. First, the two sets differed in lightness as well as hue. This was dictated by the use of the Munsell system. There are no sets of stimuli meeting the categorical and distance requirements of the experimental design at the same lightness level for the two boundaries. If the lightness levels had been the same, the conspicuousness of the boundaries may have differed. The second is a more general point. Colour categories occupy a volume in three-dimensional colour space (hue, lightness and saturation). Their boundaries are therefore surfaces, but we only sampled a point on each of the boundary surfaces, and the stimuli only varied in hue. It is therefore possible that the average degrees of conspicuousness of the two boundaries differ or that their maximum conspicuousness differs. Nevertheless, our data show that infants can detect categorical differences in the two regions by 4 months of age, consistent with hardwiring.

The lack of a perceptual distance effect is also consistent with Bornstein *et al.*. Novel same-category stimuli were looked at no more than the habituated stimulus, and there was no difference in novelty preference for the near and far new category stimuli. The lack of a within-category effect could be because the chromatic differences are below threshold, but this is very unlikely. We estimate from Knoblauch *et al.* (2001) that the chromatic JND for infants is less than 10  $\Delta E$  units. Our near and far separations were about 20 and 30  $\Delta E$ , which places them comfortably above threshold, but they did not induce increased looking. Nevertheless, one aim of the next experiment was to increase the stimulus separations further, to see if within-category novelty preference could be induced. The lack of a between-category distance effect may have been a threshold effect. Although both differences from the habituated stimulus are supra-threshold, at about 20 and 30  $\Delta E$  (as argued above), the difference between the two (10) is close to threshold. There is some suggestion in Gerhardstein *et al.*'s (1999) data, despite problems with the lighting, that the two separations are discriminably different. They used the same Munsell distances as here (three hue steps and two hue steps) and found an effect of step size. The major effect of the lighting was to differentially shrink the perceptual separations. We calculate that the perceptual distances under their lighting must have been about 20  $\Delta E$  units on average for their near pairs and 24  $\Delta E$  units on average for their far pairs (Davies & Franklin, 2002). The difference between the two is smaller than the difference here, but they still found a near-far effect.

Although our stimuli were chosen to fit the requirements of the experimental design, they are relevant to the issue of slower developmental rates for the blue channel, mentioned in the introduction (Adams *et al.*, 1994; Teller, 1998). Visual stimuli are initially coded by three different cone groups designated here as L, M and S (long, medium and short) differing in their spectral sensitivity. The L cones have their peak sensitivity at the longest wavelength, the M cones at the next longest, and the S cones at the shortest wavelengths. The cone outputs are combined into two chromatic channels that combine the cone signals in two main ways:  $L - M$  and  $S - (L + M)$ . Although these opponent channels are not sufficient to account for Herring's primaries, the  $S - (L + M)$  channel may contribute primarily to the blue-yellow axis of colour space, and the  $L - M$  cells to the red-green axis. When our stimuli were mapped onto the MacLeod-Boynton (MacLeod and Boynton, 1979) cone excitation space, it became apparent that the blue-green set were at more or less constant  $L - M$  and varied mostly in  $S - (L + M)$ . In

contrast, the blue–purple set lay diagonally in cone space, varying on both axes. Both stimulus sets produced Categorical Perception. Thus, it appears that by 4 months, the S system is sufficiently mature to produce these effects, and the size of the effects is about the same whether one channel or both channels are signalling. However, there may still be a lag in S cone development, and further maturation may yield stronger categorical effects for the primary boundary than for the secondary boundary.

## EXPERIMENT 2

### Categorical Perception for blue–purple with increased stimulus separation

The first aim of this experiment was to test the lack of within-category novelty preference further. We chose to do so by testing the blue–purple boundary again, as this was the first time that evidence for that boundary had been found in infants. Extending the stimulus separations further along the hue dimension was not possible because purple quickly becomes pink, according to adult judgments. We therefore decided to increase the separations by adding a lightness difference to the three-step hue difference used for the far pairs used in Experiment 1. Adding a lightness difference may seem perverse. Colour perception is usually deemed to be hue discrimination at constant lightness; experimenters go to great lengths to exclude lightness differences or to vary them in such a way that reliable responses could not be based on them (e.g. Knoblauch *et al.*, 2001). Indeed, criticisms of Bornstein *et al.*'s study included the possibility that infant choice was driven partly by uncontrolled brightness differences (Werner & Wooten, 1985). However, in a more general sense, colour space consists of lightness and saturation dimensions, as well as hue, and colour-order systems such as Munsell are based on that dimensional structure. Moreover, we were concerned primarily with infant colour categories, rather than establishing their sensitivities to hue alone. Linguistic categories mark volumes in colour space, and category boundaries can be crossed by changes in saturation and lightness, as well as hue. For example, in part, the difference between red and pink is due to saturation and lightness (see Experiment 3).

We chose therefore to retest the blue–purple boundary, as this was perhaps the most original finding of Experiment 1. The stimuli had the same hues as the far pairs in Experiment 1 but with two Munsell value steps added. A within-category pair differing only in value was included to see if a lightness difference alone could produce novelty preference. Deliberately including lightness variation was also pertinent to Werner and Wooten's (1985) criticism of Bornstein *et al.* They argued that uncontrolled brightness differences could have been the cause of apparent Categorical Perception. If including a lightness difference does not induce within-category novelty preference or increase between-category novelty preference, this would weaken Werner and Wooten's argument.

The methods, scoring and data analysis followed the same procedures as Experiment 1 unless otherwise stated.

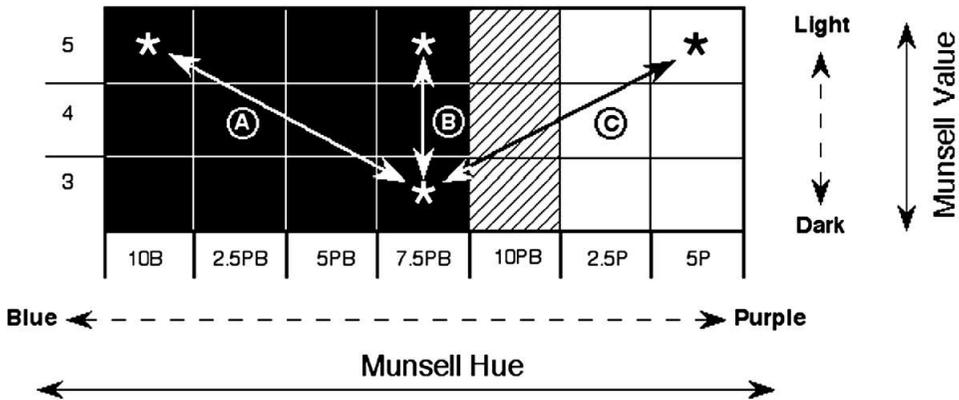
**Method**

**Participants**

Twenty-one infants were recruited from local baby clinics and groups. Three infants were excluded from the study due to general fussiness before and during the procedure. The age range of the final 18 infants was 16–20 weeks, with a mean age of 18 weeks. Eight infants were male, and 10 were female.

**Materials and stimuli**

The Munsell notations of the stimuli are given in Fig. 7. There were three stimulus pairs (A, B and C on Fig. 7). Pair A (within hue and value) and C (between hue and value) had the same Munsell hues as the blue–purple far pairs in Experiment 1 (three hue steps apart), but two value steps were added. This increased the separations by about 10 ΔE, making the separation about double those of the near pairs in Experiment 1. Pair B (value) was separated by just two steps of Munsell value with no hue separation. Six infants were randomly allocated to each experimental condition.



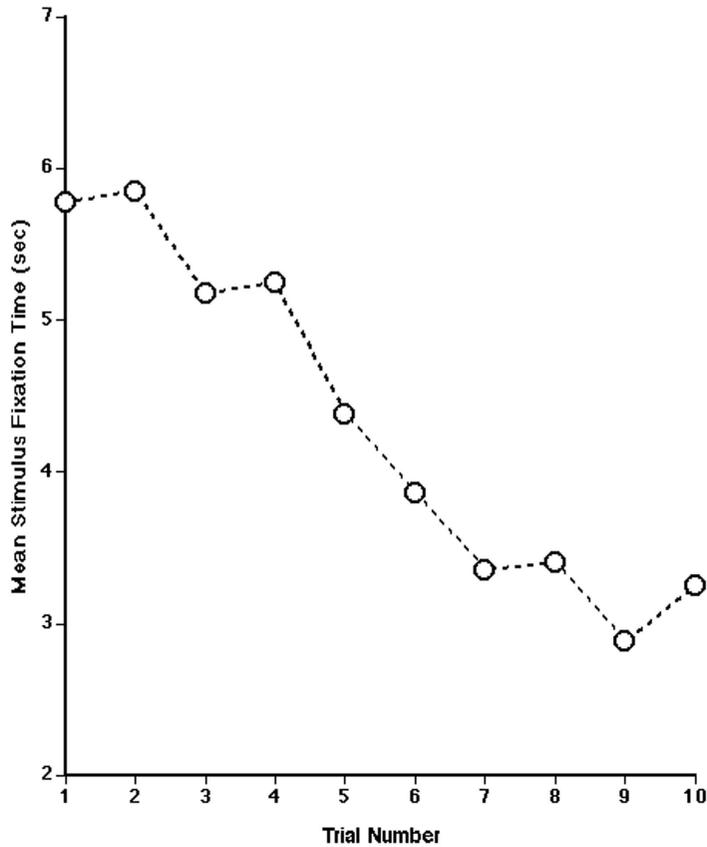
**Figure 7.** Munsell codes, categorical status and Munsell distances of the stimuli of the blue–purple hue and value experimental pairs used in Experiment 2. Pair A = within hue and value; pair B = within value; pair C = between hue and value. The shaded cell indicates an adult category boundary.

**Results**

The mean inter-observer reliability was  $r(15) = 0.84$ ,  $p < .001$ . There were no a priori preferences (largest  $F = 1.23$ , smallest  $p = 0.32$ ).

**Familiarization**

Figure 8 shows the mean stimulus fixation times over the 10 familiarization trials. Looking times on average decreased from around 6 s to around 3 s. This decreasing linear trend was significant,  $F(1, 17) = 32.23$ ,  $p < 0.001$ .



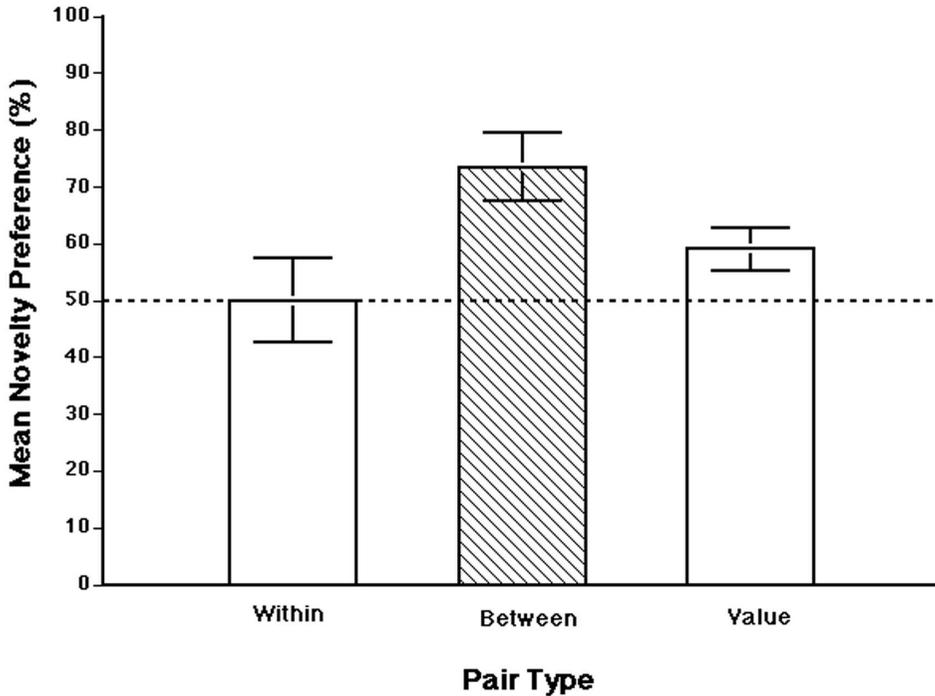
**Figure 8.** Mean stimulus fixation times during the familiarization phase for the blue–purple hue and value colour set in Experiment 2.

**Novelty preference**

Figure 9 shows the novelty preference percentages for each pair. One-sample *t* tests (two-tailed) revealed significant novelty preferences for the between-category pair,  $t(5) = 3.92, p < .05$ , but not for either within pair: hue + value  $t(5) = .12, p = .90$ ; value  $t(5) = 2.43, p = .06$ . Note, however, that the within-category value pair almost showed a significant difference, and the score of about 60% is the highest score for any within-category pair in the two experiments. The category effect was also significant (within hue + value compared with between hue + value):  $F(1, 11) = 5.519, p < .04$ . However, the two within-category pairs did not differ significantly,  $t(10) = 0.97, p = .36$ , and the difference between the between-category pair and the within-category lightness pair only approached significance,  $t(10) = 2.046, p = .068$ .

**Comparing Categorical Perception for hue with hue plus lightness**

The blue–purple far pairs used in Experiment 1 had the same hues as the current pairs but did not differ in lightness. Comparing the two novelty preference scores thus



**Figure 9.** Novelty preference for the three blue–purple experimental pairs in Experiment 2: within hue and value; between hue and value; within value.

assesses the impact of adding the lightness difference. The scores for Experiment 1 were: within = 50.46; between = 69.13. An ANOVA on the novelty preference scores with category (within/between) as a repeated-measures factor and boundary type (hue/hue + value) as an independent groups factor revealed a significant effect of category,  $F(1, 23) = 12.49$ ,  $p < .002$ , but no significant effect of boundary type or the interaction, largest  $F = .17$ , smallest  $p = 0.69$ .

## Discussion

As in Experiment 1, infants only showed novelty preference to a test stimulus from a new adult category. Thus, Categorical Perception across the blue–purple boundary was found again. The degree of novelty preference for the new category stimulus was about the same as that found in Experiment 1, even though the difference between stimuli in each pair had been increased by adding two value steps (about 10  $\Delta E$ ). Nor did adding the value steps affect within-category preference. Increasing the separation did not produce within-category novelty preference. Curiously, within-category lightness alone almost produced a significant within-category novelty preference (we return to this below). It is possible that with larger sample sizes and increased power, this difference might have been significant. However, the significant differences we have found occurred despite the small sample sizes rather than because of them. The lack of a clear lightness effect tends to support Bornstein *et al.*'s interpretation of their results and

weakens the force of Werner and Wooten's criticism, although this remark must be tempered by the relatively large within-category score for lightness alone.

The results are consistent with two main possibilities. The first is that the major determinant of preferential looking is the categorical status of the stimuli. A category shift engages the infant's attention, whereas a quantitative shift alone does not. It is possible that increasing the within-category separations further would induce novelty preference. However, this seems unlikely, bearing in mind that Bornstein *et al.* found no within-category novelty preference with much larger stimulus separations than here (see Fig. 1). Nevertheless, there is scope to increase separations further if the stimuli vary on hue, lightness and saturation. We estimate that the maximum possible within-category separation is about 60  $\Delta E$ . The between-category separation could be increased to about twice that amount. However, even if such increases produced increased novelty preference, it would not explain why we find between-category novelty preference but no within-category preference for the same separation. One possibility is that both CIE and Munsell measures of perceptual distance underestimate between-category distances. A warping factor needs to be added to the measures we have used here. We return to this in the general discussion after we provide further evidence of no within-category novelty preference.

The second possibility is that lightness and hue differences are not additive for infants. In other words, our attempt to increase the perceptual separations did not work. This is also consistent with the lack of difference between the between-category differences of the two experiments. The three dimensions of colour all contribute to perceptual distance in adults, and this is incorporated in the various colour-order systems such as Munsell and CIE as used here. Moreover, adults perceive the three dimensions 'wholistically'; in other words, the relationship among the dimensions is integral (e.g. Burns & Shepp, 1998). We defer further discussion of the latter possibility until the general discussion, as the next experiment explores the relationship among colour dimensions further.

## EXPERIMENT 3

### Categorical Perception for a lightness-saturation boundary

We argued earlier that although colour categories may be determined primarily by hue, category shifts can also occur at constant hue. The main aim of Experiment 3 was to see if Categorical Perception would be found for such boundaries. A second aim was to test another secondary boundary (pink-red) to ensure that the finding of Categorical Perception for a secondary boundary was not restricted to blue-purple. The basic requirements of the experimental design could not be met by varying value or chroma alone. We therefore used pairs varying in value and chroma (two value steps plus one chroma step) separated by about 20  $\Delta E$ . Note that this assumes that value and chroma are additive for infants.

This stimulus set was interesting for a second reason. Plotted in the MacLeod-Boynton cone excitation space, they fall at right angles to the blue-green set from Experiment 1. That is, they are signalled mainly by the L-M channel, and thus afford a test of whether categorical effects are found in this channel alone as well as in combination with S - (L + M) (blue-purple).

Methods, scoring and analysis followed the earlier procedures unless otherwise stated.

## Method

### Participants

Following Experiment 2, the same infants took part in Experiment 3.<sup>6</sup> Ten adult participants (mean age 19.6 years) from the University of Surrey (five men, five women) took part in the naming and similarity-judgments tasks. All adults had normal colour vision, as indicated by the City University Test (Fletcher, 1981).

### Materials and stimuli

The stimuli were from the pink-red region of the Munsell colour system and differed only in value and chroma. There were three pairs of stimuli: within pink, between pink-red and within red. The inter-stimulus distances were one Munsell chroma step plus two value steps (about 20  $\Delta E$ ). The Munsell notations and their relationships are given in Fig. 10. Six infants were randomly allocated to one of the three conditions.

## Results

### Adult naming and similarity judgments

The pattern of naming was completely consistent with the intended nominal category. The mean dissimilarity ratings for the pairs were: within pink = 45.7 mm ( $SD = 17.9$ ); between pink-red = 55.1 mm ( $SD = 23.6$ ); within red = 40.1 mm ( $SD = 20.1$ ). Within pairs were judged as more similar than the between pair,  $F(1, 9) = 7.50, p < .05$ .

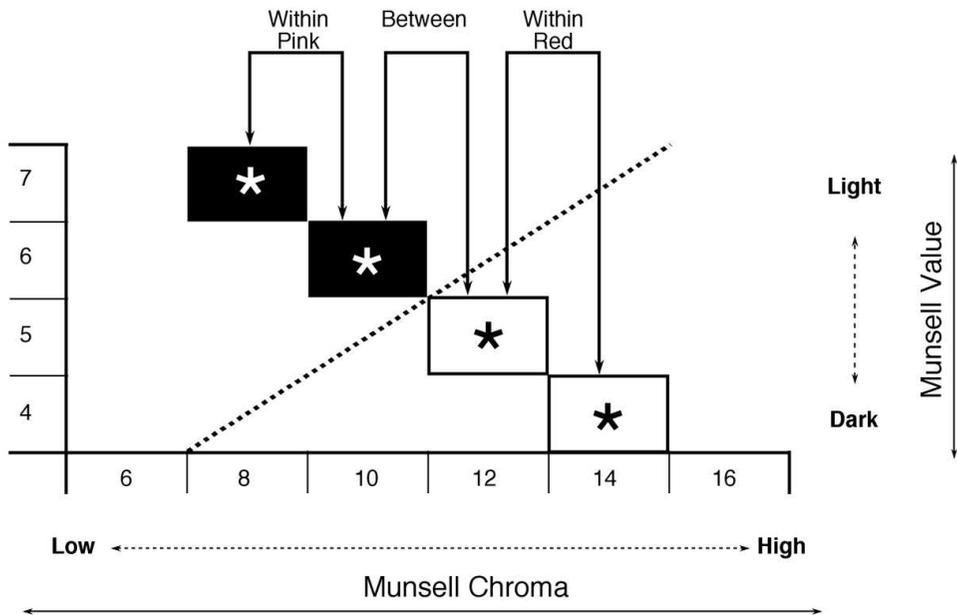
### Infant data

The mean inter-observer reliability was  $r(15) = 0.86, p < .001$ . There were no significant a priori preferences during familiarization (largest  $F = 1.85$ , smallest  $p = 0.28$ ). Figure 11 shows the mean stimulus fixation times over the 10 trials. There was a significant decreasing linear trend: Looking times on average decrease from around 5–6 s to around 3 s,  $F(1, 17) = 18.92, p < .001$ .

### Novelty preference

Figure 12 shows the novelty preference percentages for the three stimulus pairs. The pink-red pair produced novelty preference greater than chance,  $M = 70.09\%$ ;  $t(5) = 4.28, p < .01$ , but neither within-category pair did so (within pink = 45.29%; within red = 48.31%). This was supported by a one-way ANOVA on category (within/between),  $F(1, 17) = 22.79, p < .001$ .

<sup>6</sup> It is possible that infants might have looked less in Experiment 3 due to tiredness as a result of taking part in Experiment 2. However, as the results show, this does not preclude detection of novelty preference.



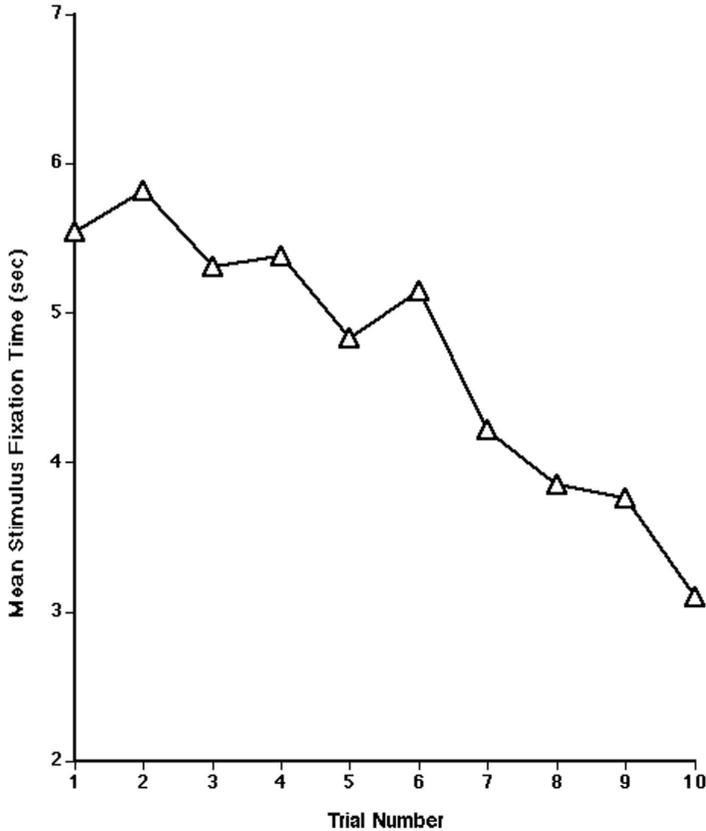
**Figure 10.** Munsell codes, categorical status and Munsell distances of the stimuli of the pink–red lightness/saturation experimental pairs used in Experiment 3. The Munsell hue is constant at 5R. The dashed line indicates the adult category boundary. The cells shaded black with a white star are the pink stimuli, and white cells with a black star are the red stimuli.

## Discussion

Infants showed novelty preference for between-category pink–red pairs and no novelty preference for equally separated within-category pairs. These results extend the evidence from Experiments 1 and 2, to imply that infants show Categorical Perception at 4 months to another secondary boundary (pink–red). Moreover, this boundary was defined by lightness and saturation rather than by hue, as in the two previous experiments, and initial stimulus coding involved only the L–M channel. Again, it appears that a category shift is required to engage the infant’s attention, rather than a quantitative change alone.

## GENERAL DISCUSSION

The results from the three experiments suggest that infants at 4 months perceive colour categorically. This was the case for the primary boundary, blue–green (as found by Bornstein *et al.*) and for two secondary boundaries: blue–purple and pink–red. The latter is particularly interesting, as it is the first time that Categorical Perception has been shown for a boundary defined by lightness and saturation rather than hue. These data are clearly inconsistent with our conjecture that primary categories are hardwired, whereas secondary categories are not. We return to the implication of this for language differences later.

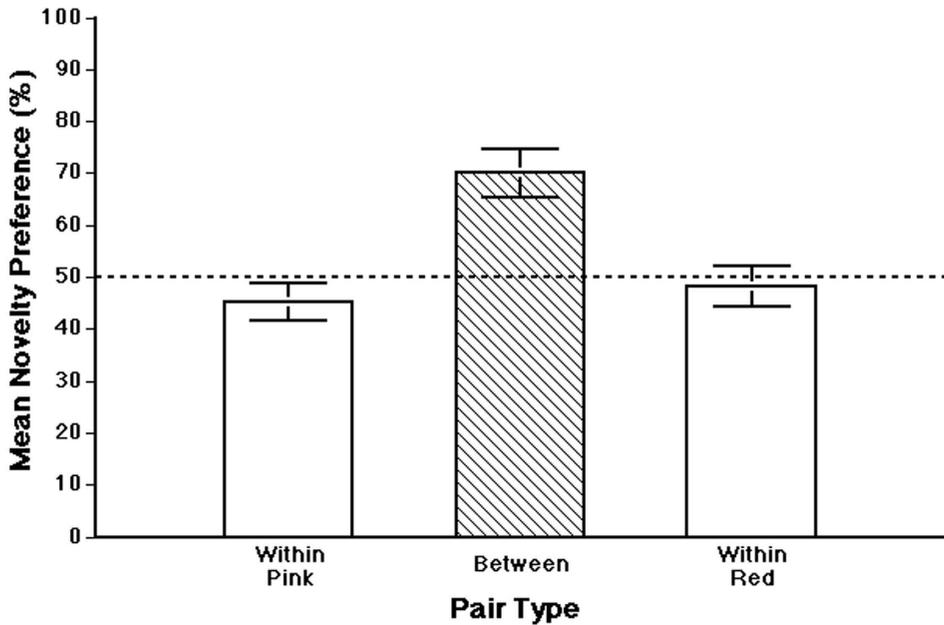


**Figure 11.** Mean stimulus fixation times during the familiarization phase for the pink-red set in Experiment 3.

Almost as consistently, no clear within-category novelty preference was found, although the stimulus separations were matched in Munsell units to the between-category pairs. The only possible exception was that the within-category lightness difference in Experiment 2 almost produced a significant novelty preference, and we return to this later. Nor did the size of the between-category separations (Experiments 1 and 2) have any effect.

The conjecture that primary categories are hardwired, whereas secondary categories are not, was intended to explain the variation in colour terms across languages. However, if the blue-purple and pink-red boundaries are as strongly hardwired as the blue-green boundary, this cannot explain the greater prevalence of languages marking the latter distinction than the other two. Similarly, assuming that these results hold universally, the relative infrequency of languages marking the blue-green boundary cannot be attributed to the lack of perceptual predisposition. Rather, it suggests that the influence of perceptual categories on linguistic categories may be relatively weak. They may be a necessary condition, but far from a sufficient condition, for linguistic category formation.

On the other hand, infant Categorical Perception for the blue-green boundary seems



**Figure 12.** Novelty preference for the three experimental pairs in Experiment 3: within red; between pink–red; within pink.

inconsistent with the absence of such effects in adult speakers of languages that do not mark the distinction (Davies, Sowden, Jerrett, Jerrett, & Corbett, 1998; Kay & Kempton, 1984; Roberson *et al.*, 2000). This apparent contradiction could be resolved in a number of ways. First, the effects we found here may not be universal. For instance, it could be that short-wavelength discrimination is restricted in speakers of 'grue' languages (i.e. languages which have one basic colour term to denote blue and green). This could be associated with high-density macula pigmentation (Bornstein, 1973), or accelerated lens brunescence (Davies, Laws, Corbett, & Jerrett, 1998) or even cone photopigment variation (Jameson, Highnote, & Wasserman, 2001). While we can rule out none of these possibilities, there is only circumstantial evidence in their favour. Second, perceptual categories may be attenuated unless reinforced by linguistic categories. There is no direct evidence for such category loss. However, there is evidence for selective changes in discriminability as a function of position within the category. There is a parallel for this in the perception of speech sounds. Language learning appears to reduce within-phoneme discriminability (Werker & Tees, 1984). There is also evidence that category learning produces 'learned equivalence' (reduced within-category discrimination) (Goldstone, 1994; Özgen & Davies, 2002). The blue-green boundary falls in the centre of grue categories and may therefore be subject to maximum learned equivalence. There is no direct evidence for this, other than the lack of a Categorical Perception effect, but, in principle, it is easily testable. In addition, new colour categories can be learned, and the learning produces enhanced between-category discrimination. Taken together (learned equivalence and learned distinctiveness), it seems plausible that hardwired Categorical Perception may be lost if the boundary falls in the centre of a linguistic category. Third, it could be that infant

Categorical Perception, as evidenced by preferential looking, and adult Categorical Perception, as evidenced by measures such as RTs in same-different tasks, similarity judgments, triadic judgments and successive recognition (2-AFC) tasks are not the same phenomena. The various adult measures converge, to a degree at least, on a common explanation, but there is no corresponding variety of infant measures. This deficiency needs to be addressed and could be done by broadening the range of measures and basic tasks used. For instance, Gerhardstein *et al.* (1999) used a method where the infant could control what it could see, and the task was predicated on the idea of attention-free detection (pop-out). If infants showed categorical effects in this task (and others), it would tend to support the equivalence of infant and adult Categorical Perception.

We suggested, tentatively, in the discussion of Experiment 2 that hue and lightness might not be additive in infants. This was because adding a lightness difference to the hue separations of Experiment 1 did not either produce within-category novelty preference or increase the size of the between-category score. This pattern is also consistent with novelty preference being determined by an all-or-none categorical change. However, Experiment 2 also provided the sole exception to an otherwise consistent pattern across the experiments. The within-category novelty preference score for the pair differing in lightness produced the largest of the within-category scores (about 60%), although this was not quite significant. If we assume that lightness and hue differences are not additive for infants, and that if both vary, infants only attend to hue, this is consistent with the data. Increasing the stimulus separations from Experiment 1 was ineffective because only the hue difference was operative. However, when only lightness varied, this could have determined infants' within-category preference. The weakness of this account is that it does not explain why a 10  $\Delta E$  lightness difference induces novelty preference, whereas a 20  $\Delta E$  hue difference does not. The stimulus separations in Experiment 3 combined value and chroma, with each component contributing about 10  $\Delta E$ . This produced about the same size between-category effect as in Experiments 1 and 2, and is consistent with the dimensions being additive. However, we cannot rule out the possibility that one or other of the components was dominant. Given that the integrality of hue, lightness and saturation is well established for adults (e.g. Burns & Shepp, 1998), and there is no direct evidence that they are not in infants, it is more parsimonious to assume that the within-category lightness data from Experiment 2 are 'noise'. However, if the result proved to be reliable, it would be worth investigating the relationship among colour dimensions further.

Finally, we note that categorical effects were found based on initial coding by the S - (L + M) channel (blue-green) and by the L-M channel (red-pink) and by their combination (blue-purple). These data argue against a possible source of the categorical effects in early visual processes. It was suggested that within-category coding was based primarily on single chromatic channel, S - (L +) or L-M, and that category boundaries might form at the transition between coding by one channel to coding by the other. The involvement of both channels in between-category discrimination could be the origin of perceptual warping. However, our earlier remarks about the blue-green and red-pink set appear to rule out this explanation.

## Conclusion

Our data imply that 4-month-old infants seem to have adult perceptual colour categories, at least to some degree. The evidence is equally strong for a primary boundary (blue-green) and two secondary boundaries (blue-purple and pink-red). Thus, these data offer no support to the conjecture that the Berlin and Kay hierarchy of basic colour terms reflects differing degrees of salience of hard-wired universal categories. For these infants, the influences of categorical separation and perceptual distance are not additive, at least under the measurement operation used here.

## Acknowledgements

This research was supported in part by Economic and Social Research Council Postgraduate Studentship R42200124191 to the first author and in part by Economic and Social Research Council Grant R000223894 to Jules Davidoff, Debi Roberson, Ian Davies and Michelle de Haan. We are grateful to Sarah Coughtrey, Ally Clifford, Amy Riddett and Claire Campbell for assistance with the some of the data collection. We thank Nigel Woodger for designing and constructing the babylab apparatus set-up, and for the design of the publicity materials. We also owe thanks to Sam Boyles, Emre Özgen and Paul Sowden for constructive comments on drafts of this paper. We are also grateful to three reviewers of this paper.

## References

- Adams, R. J., Courage, M. L., & Mercer, M. E. (1994). Systematic measurement of human neonatal color vision. *Vision Research*, *34*(13), 1691-1701.
- Berlin, B., & Kay, P. (1969). *Basic color terms: Their universality and evolution*. Berkeley: University of California Press.
- Bornstein, M. H. (1973). Color vision and color naming: A psychophysiological hypothesis of cultural difference. *Psychological Bulletin*, *80*, 257-285.
- Bornstein, M. H., Kessen, W., & Weisskopf, 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. O. (1984). Discrimination and matching within and between hues measured by reaction times: Some implications for Categorical Perception and levels of information processing. *Psychological Research*, *46*, 207-222.
- Boynton, R. M., Fargo, L., Olson, C. X., & Smallman, H. S. (1989). Category effects in color memory. *Color Research and Application*, *14*, 229-234.
- Burns, B., & Shepp, B. (1998) Dimensional interactions and the structure of psychological space: The representation of hue, saturation and brightness. *Perception and Psychophysics*, *43*(5), 494-507.
- Catherwood, D., Crassini, B. & Freiberg, K. (1990). The course of infant memory for hue. *Australian Journal of Psychology*, *42*, 277-285.
- Davies, I., & Franklin, A. (2002). Categorical similarity may affect colour pop-out in infants after all. *British Journal of Developmental Psychology*, *20*, 185-203.
- Davies, I. R. L., Laws, G., Corbett, G. G., & Jerrett, D. J., (1998). Cross-cultural differences in colour vision: Acquired 'colour-blindness' in Africa. *Personality and Individual Differences*, *25*, 1153-1162.
- Davies, I. R. L., Sowden, P., Jerrett, D. T., Jerrett, T., & Corbett, C. G. (1998). A cross-cultural study

- of English and Setswana speakers on a colour triads task: A test of the Sapir-Whorf hypothesis. *British Journal of Psychology*, *89*, 1-15.
- De Valois, R. L., & De Valois, K. K. (1990). *Neural coding of color. Readings on color, vol 2: The science of color* (pp. 93-140). Cambridge, MA: The MIT Press.
- De Valois, R. L., & Jacobs, G. H. (1968). Primate color vision. *Science*, *162*(2853), 533-540.
- Fagan, J. (1974). Infant color perception. *Science*, *183*, 973-975.
- Fahle, M., & Poggio, T. (Eds.). (2002). *Perceptual learning*. Cambridge, MA: MIT Press.
- Fletcher, R. (1981). *City University Colour Vision Test* (2nd ed.). Windsor, UK: Keeler.
- Gerhardstein, P., Renner, P., & Rovee-Collier, C. (1999). The roles of perceptual and categorical similarity in colour pop-out in infants. *British Journal of Developmental Psychology*, *17*, 403-420.
- Goldstone, R. L. (1994). Influences of categorisation on perceptual discrimination. *Journal of Experimental Psychology: General*, *123*(2), 178-200.
- Harnad, S. (1987). Psychophysical and cognitive aspects of Categorical Perception: A critical overview. In S. Harnad (Ed.), *Categorical Perception: The groundwork of cognition* (pp. 535-565). New York: Cambridge University Press.
- Hering, E. (1878). *Outlines of a theory of the light sense*. Translated from German by L. Hurvich & D. Jameson, 1964. Cambridge, MA: Harvard University Press.
- Hunt, R. W. G. (1987). *Measuring colour*. Chichester, UK: Horwood.
- 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., Berlin, B., & Merrifield, W. R. (1991). Biocultural implications of systems of color naming. *Journal of Linguistic Anthropology*, *1*, 12-25.
- Kay, P., & Kempton, W. (1984). What is the Sapir-Whorf hypothesis? *American Anthropologist*, *86*, 65-79.
- Kay, P., & Maffi, L. (1999). Color appearance and the emergence and evolution of basic color lexicons. *American Anthropologist*, *101*, 743-760.
- Kay, P., & McDaniel, C. K. (1978). The linguistic significance of the meanings of basic color terms. *Language*, *54*, 610-646.
- Knoblauch, K., Vital-Durand, F., & Barbur, J. L. (2001). Variation of chromatic sensitivity across the life span. *Vision Research*, *41*, 23-36.
- Krauskopf, J., Willams, D. R., & Heeley, D. W. (1982). Cardinal directions in color space. *Vision Research*, *22*, 1123-1131.
- MacLeod, D. I. A., & Boynton, R. M. (1979). Chromaticity diagram showing cone excitation by stimuli of equal luminance. *Journal of the Optical Society of America*, *69*, 1183-1186.
- 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.
- Pereverzeva, M., Chien, S. H-L., Palmer, J., & Teller, D. Y. (2002). Infant photometry: Are mean adult isoluminance values a sufficient approximation to individual infant values? *Vision Research*, *42*, 1639-1649.
- Pilling, M., & Davies, I. R. L. (in press) Linguistic relativism and colour cognition. *British Journal of Psychology*.
- Roberson, D., & Davidoff, J. (2000). Categorical Perception of colors and facial expressions: The effect of verbal interference. *Memory and Cognition*, *28*, 977-986.
- Roberson, D., Davidoff, J., & Davies, I. R. L. (2000). Color categories are not universal: Replications and new evidence from a Stone-Age culture. *Journal of Experimental Psychology: General*, *129*, 369-398.
- Teller, D. Y. (1998). Spatial and temporal aspects of infant color vision. *Vision Research*, *38*, 3275-3282.
- 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.

- Werner, J. S., & Wooten, B. R. (1985). Unsettled issues in infant colour development. *Infant Behavior and Development*, 8, 99-107.
- Wooten, B. R. (1974). Infant hue discrimination? *Science*, 187, 275-277.
- Wyszecki, G., & Stiles, W. S. (2000). *Color science: Concepts and methods, quantitative data and formulae*. New York: Wiley.

Received 3 June 2003; revised version received 10 July 2003