Colour Categorisation in Infants

Background

Theoretical Positions: Universal vs Linguistically Relative Colour Categories

The physiological basis of colour vision available to all humans with normal trichromatic colour vision (Jordan & Mollon, 1997) allows discrimination between millions of different colours. Yet there is a considerable inclination to group these colours into a relatively small number of categories. One only has to consider the acronymic mantra "Richard Of York Gained Battles In Vain" chanted at school in honour of Newton who, for reasons of alchemy, declared 7 colours in the spectrum; or look at children's drawings of rainbows to show the power of these categorical structures. However, while there is no doubt that we can perceive colours categorically, there has been considerable debate as to whether the origin of these categories is innate (Universal) or derived from labels in the speaker's language (Berlin & Kay, 1969; Bornstein, Kessen & Weiskopf, 1976; Heider & Olivier, 1972; Roberson, Davies & Davidoff, 2000; Saunders & van Brakel, 1997; Stefflre, Castillo & Morley, 1966).

A powerful argument in favour of innate -and hence in favour of the universal colour categories proposed by Berlin and Kay (1969)- comes from the Bornstein et al. (1976) study of 4-month-old infants. In that study, infants showed categorical perception for all primary colours. However, the intuitive appeal of the data is based on an implicit false premise. Kay and McDaniel (1978) proposed that primary colours (red, green, yellow and blue) derive from the output of opponent-process cells but this is a mistake (De Valois & De Valois, 1990). The error comes from the shorthand of calling these cells red/green and yellow/blue. Opponent-process cells do not output categorical information; after habituation for light of say a long wavelength ("red"), they will simply output more in the middle part of the visible spectrum ("green"). The combined output of opponent-process cells does, in fact, produce wavelengths at which discrimination is clearly more sensitive. These discontinuities might have produced colour categories but the wavelengths are not at the required places to get primary colour categories. Nor do colours that we see as uniquely red, yellow, green or blue have any clear relation to these discontinuities (Saunders & van Brakel, 1997). So, to make the point clear, there is no widely accepted neurophysiological evidence that would make colour categories inevitable.

Bornstein (1985) recognised that some difficulties arise from the proposal of innate colour categories. The first is explaining the well-known difficulty children have in learning colour names. If the physiological apparatus is already in place at 4 months, it seems odd that it takes another 18 months to learn the first colour word during the time the child shows a spurt of word learning. Bornstein (1985) suggested that there must be some maturational delay perhaps of callosal fibres; a proposal that does not explain the dramatic increase in the acquisition of colour terms of present day 4-year-olds compared to those of 100 years ago. Bornstein (1985) was also aware that a colour vision system in line with Western colour terms would, if innate, pose problems for the many languages in the world that did not possess these terms. Children in those languages ought to find colour naming even more difficult. Our recent ESRC funded project (R0002383-10) has investigated Bornstein's conjecture in a 5-term language in Namibia. It was found that those children were not disadvantaged. Moreover, children who knew no colour names, in Namibia and the UK, showed errors based purely on perceptual distance (Roberson et al. submitted) and not related to Berlin and Kay's universal colours

Aims of the Proposal

Bornstein (1975; Bornstein et al, 1976) produced more than one result in favour of innate colour categories. The most well-known is the classic study with babies (Bornstein et al., 1976) showing novelty preference for a stimulus from a different colour category but not from the same category. However, he also showed that given a choice between a colour from the centre of a category or from the boundary of the category, infants looked longer at the central (typical) colour (Bornstein, 1975). Doubts were raised at the time about the methodology in Bornstein's studies (Banks & Salapatek, 1983; Werner & Wooten, 1985) but there have never been replications. The purpose of our project was to consider two of his lines of evidence in better-controlled studies.

It soon became apparent that it was not going to be easy to complete all the experiments that were planned in our proposal. There was a minor delay in getting the software originally designed for two monitors to work with three. As we wanted to minimise
the role of the experimenter in our studies and allow automatic presentation of accurately

defined stimuli, it was important to get that right. However, the longer delay in getting started was from our initial intention to use 2 Munsell step differences between stimuli as in Bornstein and Korda (1984, 1985). Note that we use here the nomenclature of Bornstein and Korda (1984). The steps really refer to the commercially available paper samples; with our computer produced stimuli, 25 gradations can be produced within each of the commercially available hue steps. It became clear after a few months that our babies were showing no novelty preference for these 2-step differences. We therefore decided to increase the step size using the blue/green boundary because somewhat close to half of all possible colours are contained within that region and it allows the maximum range of within-category and cross-category alternatives.

We conducted a pilot experiment with adults in the two-alternative forced-choice procedure developed by Etcoff and Magee (1992) and used extensively since (see Roberson et al, 2000). The pilot established robust cross-category advantages for stimuli at 4 step differences. Having obtained the maximum possible within-category distances at that saturation, we restarted the recruitment of babies but the delay left only 6 months available for data collection. However, we were fortunate that one of the co-applicants (Ian Davies) had started a Ph.D student investigating problems within Bornstein's data. Their data were not being collected using computer produced stimuli but we imagined that they would give similar, though perhaps noisier, findings. In that belief, we were shown to be wrong and this report aims to present both sets of results and provide some potential explanations of the different findings.

Results

Our original plan was to conduct two experiments. The first was to consider Bornstein's claim, driven by the Universal position, that babies were more interested in looking at typical rather than atypical colours. Our aim was to do this with a simple preferential looking procedure with saturation and brightness controlled. While the project was unable to find time to carry out this -the less important of the two experiments- relevant data are in Franklin and Davies (in press). The relevant data come from their familiarisation procedure, similar to that used by Bornstein, in which the baby looks a decreasing amount of time at the stimulus in successive fixed periods (trials). Franklin and Davies were not explicitly concerned with the issue of typicality and it is only fortuitous that one of their comparisons was between typical and atypical purple. Fig 1 shows the amount of time the babies spent looking at the two stimuli which is clearly similar t (7) < 1. So, though our examination of hue preferences proposed by Bornstein (1975) is restricted to one colour category, we find when saturation is controlled no evidence that the 4-month-old infant prefers to look at a typical colour.

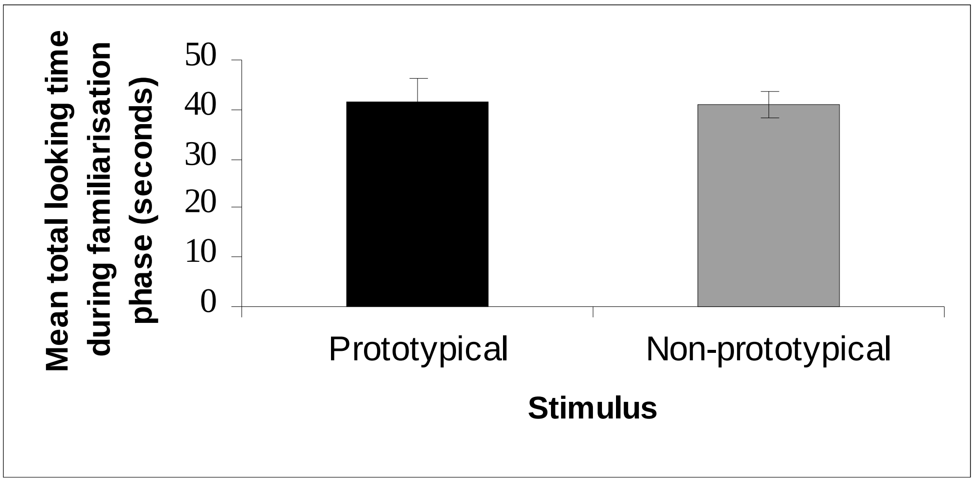

Figure 1 Mean total looking time during the familiarisation phase for the prototypical purple (2.5P 3/10) and the non-prototypical purple (7.5P 3/10).

The major experiment proposed was an examination of novelty preference within and between categories after stimulus distance had been properly controlled. Three different comparisons were envisaged: blue vs green, blue vs purple and an unbalanced blue vs green comparison used in Sandell, Gross and Bornstein (1979). As it became clear that preferences were emerging in the Franklin and Davies study and none with the Proposal study, it was decided to concentrate on increasing the power of the most important comparison: blue vs green.

The final sample for the Proposal study was 22 babies who all habituated to a green stimulus; the number for comparable conditions in Experiment 1 of Franklin and Davies was 18. The Proposal study used a within-participant design; each baby being able to offer novelty preference for a within-category (green) and a between-category alternative (blue). The Franklin and Davies study was divided into 6 conditions with 6 babies/condition. Three of these conditions concerned the blue-green category in a between-participant design. Six babies were in the within-category condition and 6 each in a near (2step) and far (3step) cross-category condition. The between-category condition of the Proposal study was 4step. The data from the Proposal study are given in Fig 2 and those from Franklin and Davies in Fig 3. Franklin and Davies show novelty preference across blue/green categories; the Proposal study did not despite having more babies for this specific comparison and using the more powerful within-participants design. The Franklin and Davies study shows preference for the novel blue stimulus to the same extent as the Proposal study shows for the novel red alternative given as a marker of habituation at the end of the test trials.

There are two other interesting aspects of the data in Fig 2 and Fig3. First, there is chance preference for novel within category choices in both studies. Given that 4-month-old babies are reported to have the same shape spectral sensitivity function, even if not the same sensitivity, as adults (Pereverzeva, Chien, Palmer & Teller, 2002) and the difference between stimuli is above threshold (Knoblauch, Vital-Durand & Barbur, 2001), the lack of within category novelty preference may seem surprising. However, Bornstein et al. also found no within-category effects. Second, in the Franklin and Davies study where cross category differences were found, there was no difference between near and far cross-category comparisons. Those data would seem to imply innate differences that are predominantly sensitive to boundary changes. In that respect, they do not support the latest version of the Universal position (Kay & Regier, 2003) in which it is allowed that boundary differences could result from linguistic differences while maintaining the important of category focals.

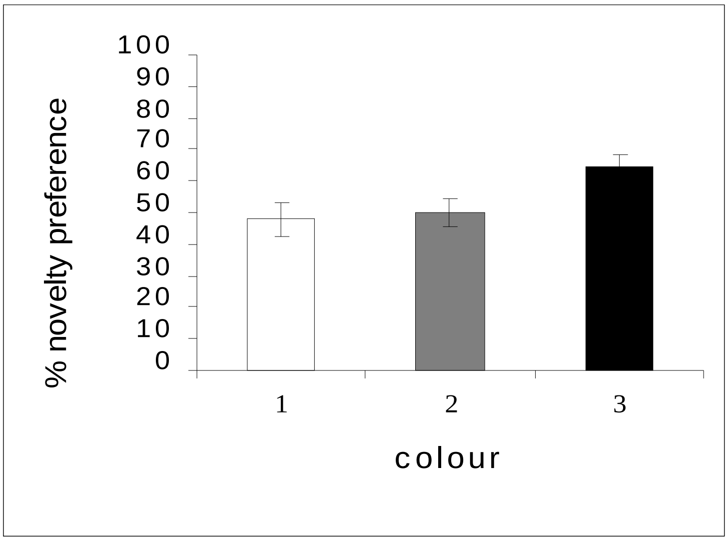

Figure 2. In the Proposal study, novelty preference is only shown for the post-test check

comparison to red (3). There is no novelty preference for within green (1) or cross blue (2). Chance is at 50%

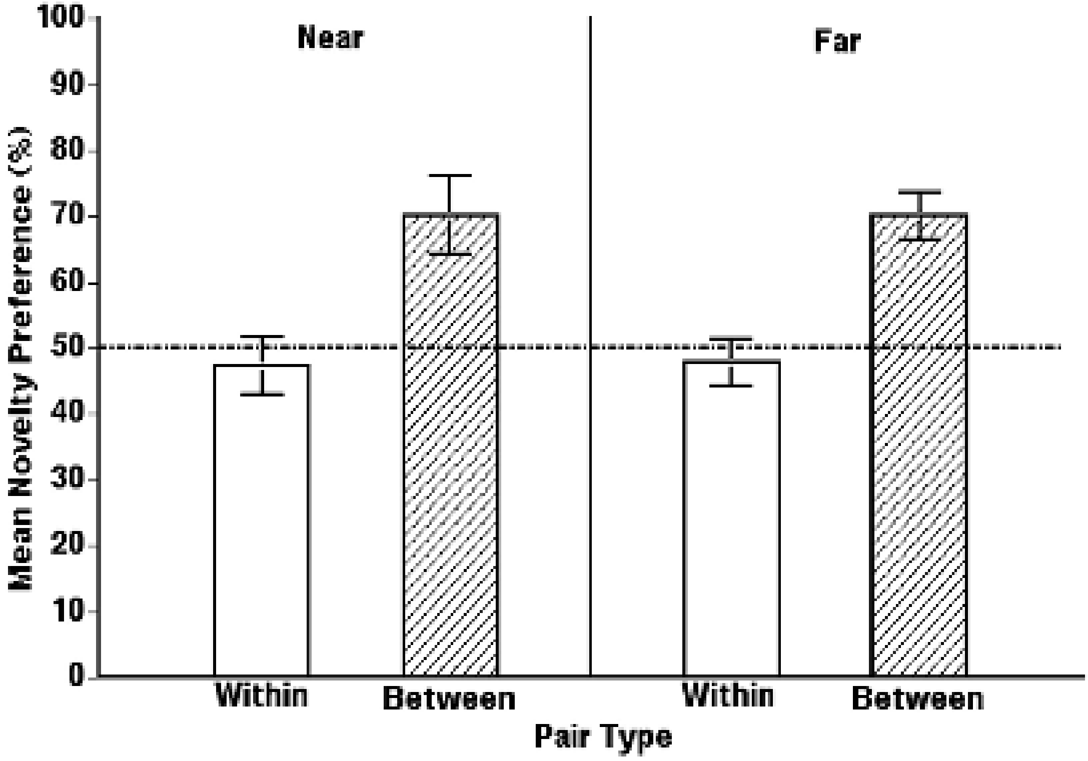

Figure 3. In Franklin & Davies, novelty preference is shown across categories (blue vs green) for both near (2 step) and far (3step) distances. There is no novelty preference for these distances within categories

As Franklin and Davies point out, their lack of within-category novelty preference would argue for equivalence within categories of very much the same type as found for phoneme discriminability in young infants (Werker & Tees, 1984). However, it is not reasonable to argue that the equivalence is learned as it can be for phoneme discrimination. If they are innate, such marked categorical effects cannot be the result of retinal processing. One would need to argue for innate colour rather than wavelength coded cells within the visual cortex for which there is no support. Nevertheless, Yoshioko, Dow and Vautin (1996) have reported cells that fire selectively for combinations of wavelength and brightness. Such combinations would be essential if there were to be category coded cells for say pink as found in Franklin and Davies (in press). In their study, both primary (blue, red) and secondary colours (purple, pink) would appear to sharply separated in the four-month old with little or no interest in making distinctions within these categories. The conclusions could hardly be more different to those from our cross-lingual studies of children who know no colour names (Roberson et al, submitted). We now consider the similarities and differences between the Proposal study and the equivalent experiments in Franklin and Davies.

Comparison of Proposal and Franklin & Davies

<u>Similarities</u>

The basic design and procedures of both studies were similar (see Table 1). Bornstein et al (1976) used a release from habituation by presenting a single novel stimulus. Both studies departed from Bornstein et al (1976) by measuring a novelty preference in a paired comparisons procedure in which both the familiar and novel stimuli were presented simultaneously following familiarisation. Mean total habituation time for green in the Proposal study was 44.12 secs similar to the 35 secs for blue/green of Franklin and Davies. The stimuli were rather larger in the Proposal study but the baby placed a little further away to give very much the same visual angle for both studies.

Data were coded blind in both studies though the manual changing in Franklin and Davies would inevitably mean that their experimenter would know the identity of the stimuli; however, as the baby could not see the experimenters, the chance of inadvertent cueing of the babies can be ruled out. Observer reliability was also high in both studies. For the Proposal study, it was .71 for green looking times and .65 for blue; while both highly significant ($p < .005$) these are lower reliabilities than those in Franklin and Davies. However, the reliability measures in the Proposal study were for all observations and not just for a small random sample. The looking times in the Proposal study were the mean of the two observers for every observation in each of the conditions. Notwithstanding the similarities, there were differences between the studies and these are summarised in Table 2. Inspection of Figs 2 and 3 might lead to the conclusion that the Proposal procedures had not produced green/blue category effects because they were less sensitive. The major point in discussing the differences between the studies is to examine that possibility.

<u>Differences</u>

One of the problems of using spectral stimuli, as in Bornstein's studies, is that saturation cannot be equated across colour categories. Saturation could well be a factor in attracting attention to coloured stimuli. As can be seen in Fig 4, Franklin and Davies used higher saturated stimuli than those in the Proposal study. These would be a little nearer Bornstein's stimuli of coloured lights that would be located right against the boundary of the CIE diagram. However, a priori, it would seem unlikely that the small differences in saturation would produce such a marked difference between the studies. There might also have been greater contrast for the Franklin and Davies stimuli that had black surrounds. The Proposal colours had a grey surround on the monitors though the set of monitors had a black surround.

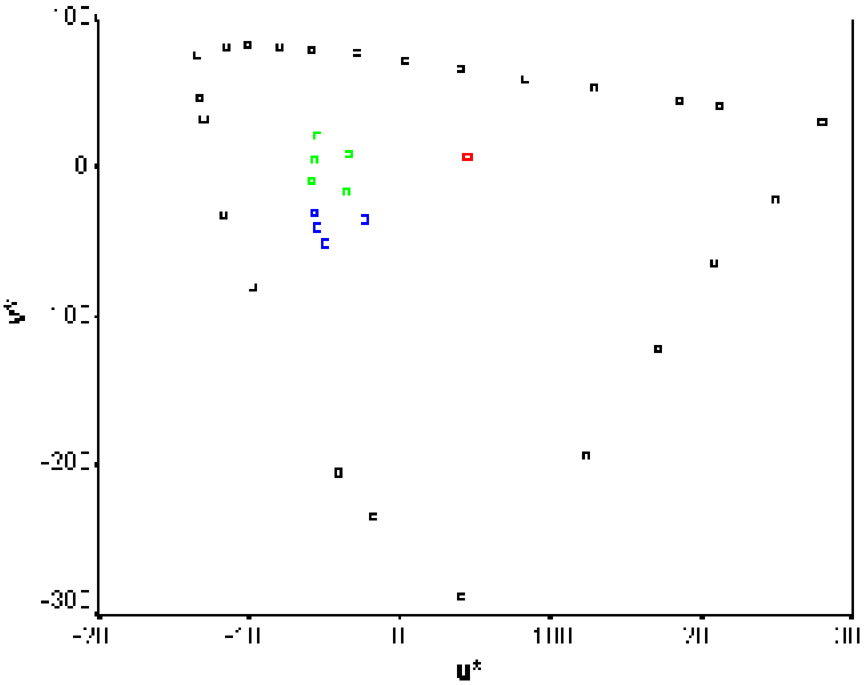

Figure 4. The position of the colours used in the Proposal study (10G5/6, 10BG5/6, 10B5/6) and in Franklin & Davies (7.5G5/10, 2.5BG5/10, 5BG5/10, 10BG5/10, 2.5B5/10, 5B5/10) set within the boundary of the u* v* CIE diagram of possible colours. The green to blue colours are arranged vertically with the Proposal study colours closer to the centre of the diagram. The post test red (3R5/6) in the Proposal study is at the other side of the white point (0,0).

In considering which aspects of the stimuli create novelty preference, as Franklin and Davies suggest, it could be that one should simply use perceptual distance (combination of hue, brightness and saturation) for within-category comparisons. Thus, novel brightness differences alone would not attract the babies' attention unless they alone would produce the critical difference in perceptual distance. The procedure would imply that Bornstein's efforts, and those of all investigators since to maintain equiluminance were not of the first importance. There was a trend in Franklin and Davies Experiment 2 ($p = .06$) for a novelty effect of brightness alone but this, to be compatible with the rest of the data, needs to be seen as noise.

Considering stimulus differences in terms of perceptual distance rather than Munsell step sizes does have consequences for estimating how far apart were the Proposal study stimuli. Due to their lower saturation, they are not as far apart as the 3-step stimuli of Franklin and Davies. However, that separation was sufficient to produce categorical effects in Franklin and Davies but not in the Proposal study. So, it is unlikely that stimulus characteristics can explain the lack of categorical effects in the Proposal study.

For familiarisation, the Proposal chose habituation-to-criterion, rather than fixed length trials as used by Bornstein because it is considered to be a 'stronger' learning procedure. The habituation continued until the baby had two consecutive looks whose average length was less than 50% of the average of their two longest looks (this is called a 'floating' criterion). It guarantees that each baby has looked long enough to decrease their interest substantially. With a fixed familiarisation, it is possible that you end up with no habituation even for the whole group. Moreover, even if there is a significant effect for the whole group, you do not know that each baby actually decreased looking time. The Proposal used this stronger procedure to avoid potential problems/criticism if we did fail to find novelty preferences, so that this failure could not be blamed on inadequate exposure during the learning phase.

Franklin and Davies presented two identical stimuli at familiarisation instead of one centrally placed. It is possible that this would encourage the baby to view both stimuli in the test phase but in case there was a bias their data were weighted by side preference. The Proposal study eliminated babies with marked side preferences though there were very few. The weighting made no difference to the Franklin and Davies results because the strong category effects are present in the raw data $F(1,20) = 19.29$, $p < .0001$. However, there is one aspect of the difference between the procedures that could be important.

Babies not only habituate to particular objects, they can also habituate to location. In the well-established procedure used in the Proposal, the baby shows novelty preference in paired comparison tests following habituation to a single central stimulus at a different location (Slater, Morison & Rose, 1982). The procedure has an advantage for young infants because, at 3 months, novel locations are as interesting as novel objects, and only at 6 months do novel objects become more interesting than novel locations (Harman et al, 1994). However, due to an inability to program eye movements to attended locations, three-month-olds show inhibition of return for 10 degrees target eccentricities, but not for 30 degrees target eccentricities. It could be that familiarisation to two different locations at 17 deg separation makes the 4-month-old babies in Franklin and Davies more likely to keep their attention at these locations. Alternatively, in the Proposal study, the conjoint change in colour *and* location produced dishabituation to location that may have diluted or overridden colour dishabituation.

Whatever the differences between the familiarisation procedures, could it be that the babies were insufficiently habituated in the Proposal study? If so, it was not because the babies had less familiarisation experience (see Table 1). There was also no trend for a greater category novelty preference if the baby habituated for a longer time ($r = -.127$). Could it be that the category novelty preference dissipated quickly in the Proposal study? We examined the category novelty preference for the first test presentations of the novel blue and green but it remained stubbornly absent ($t(21) < 1$). Could it be that there is some inadequacy about the procedure for assessing habituation in the Proposal study? The babies were tested for red preference after the green-blue category test trials and were only included if showing that preference. The novelty preference for red in the included sample was markedly greater than 50% ($t(21) = 5.86$, $p < .001$). So, to some extent they must have been habituated to the original green stimulus. One unexcluded possibility is that babies had some innate preference for red; this is perhaps unlikely though not specifically ruled out. The one possible indication of inadequate habituation came from a consideration of whether there was any greater category novelty preference for those babies who showed more red preference. Here there might be a trend ($p = .093$). However, the trend was not greater in the

12 babies who showed preferences greater than 60%. Perhaps the trend should be regarded as noise, perhaps not.

One important difference between the studies- in fact, a large motivation for the project- was to introduce computer controlled stimulus presentations. Change of stimuli in the test phase was noiseless because computer controlled; it is not clear whether that was the case for Bornstein et. al but was not so for Franklin and Davies because stimuli were changed manually by an assistant in the room. Clearly it is paradoxical, that the low-technology solution produced a greater sensitivity to categorical perception.

Table 1 Similarities between Studies

	Proposal	Franklin & Davies
Mean age	18.7 weeks	17 weeks
St dimensions	18 deg	17 deg
St separation	18 deg	17 deg
Total Familiarisation Time	44.12 secs	35 secs
Test procedure	Preferential looking	Preferential looking
Test	4, 10sec trials	5, 8 second trials used (Bornstein 9, 15 second trials)
Coding	Blind	Blind

Table 2

Differences Between Studies

	Proposal	Franklin & Davies
Design	Within-participants	Between-participants

Stimuli	Chroma 6	Chroma 10
Familiarisation	To criterion	Fixed number of trials (10)
Familiarisation	Central	Two identical stimuli 17 deg apart
Post-test check	Novelty to red	None
Stimuli	On monitor	Paper
Presentation	Computerised	Manually change
Drop-out rate	14/36	19/55

Activities

Output

Franklin, A. & Davies, I.R.L. (in press) New evidence for infant colour categories . British Journal of Developmental Psychology

Future Research

The Project has raised more questions than it solved. The replication of Bornstein's work with better control over stimulus differences resulted in the failure to find colour categories in the 4-month-old infant. We also failed to find any evidence for Bornstein's claim that the infant prefers typical colours. However, drawing conclusions from the null results of the Proposal study are difficult given the clear categorical differences, independent of lightness and other perceptual distance constraints, found in Franklin and Davies. The lack of resolution from the project does not diminish the importance of the question being tackled. It is of critical importance to know whether colour categories have, or do not have, an innate basis and future research must resolve the matter.

Future research besides giving a more definitive answer to the nature/nurture issue needs to consider the now often repeated lack of within-category preference in the 4-month-old infant. If we are to believe that the stimuli in our studies are all above threshold (Pereverzeva et al, 2002; Knoblauch et al, 2001), it raises serious questions about what is being measured by novelty preference. Future research will need to use a range of converging methods to give answers to these questions. It could well benefit from combining behavioural research with techniques from neuroimaging.

References

Banks, M.S.& Salapatek, P. (1983) Infant visual perception. In M.M.Haith and J.J.Campos (Eds) Handbook of Child Psychology: Vol 2. Infancy and Developmental Psychobiology. New York: Wiley.

Berlin, B. & Kay, P. (1969) Basic color terms: Their universality and evolution. Berkeley: University of California Press.

Bornstein, M.H. (1975). Qualities of color vision in infancy. Journal of Experimental Child Psychology, 19, 401-419.

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

Bornstein, M.H., Kessen, W. & Weiskopf, S. (1976). Color vision and hue categorization in young human infants. Journal of Experimental Psychology: Human Perception and Performance, 2, 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.

Bornstein, M.H. and Korda , N.O. (1985). Identification and adaptation of hue: Parallels in the operation of mechanisms that underlie categorical perception in vision and in audition. Psychological Research, 47, 1-17.

De ValoisR.L. & De Valois, K.K. (1990). Neural coding of color. Readings on color Vol 2: The science of color (pp93-140). Cambridge MA: MIT Press

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

Franklin, A. & Davies, I.R.L. (in press) New evidence for infant colour categories . British Journal of Developmental Psychology

Harman C, Posner, M.I., Rothbart, M.K. & Thomas-Thrapp L.(1994) Development of orienting to locations and objects in human infants.Canadian Journal of Experimental Psychology,301-18.

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

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

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

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

Knoblauch, K., Vital-Durand, F & Barbur, J.L. (2001). Variation in chromatic sensitivity across the life span Vision Research, 41, 23-36.

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.

Roberson, D., Davidoff, J., Davies, I. & Shapiro, L. (submitted) Color categories: Evidence for the relativity hypothesis.

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

Sandell, J.H., Gross, C.G. & Bornstein, M.H. (1979). Color categories in macaques. Journal of Comparative and Physiological Psychology, 93, 626-635.

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

Slater A, Morison V, Rose D. Visual memory at birth.(1982) British Journal of Psychology, 73, 519-25

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

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.

Yoshioko, T., Dow, B.M. & Vautin, R.G. (1996). Neural mechanisms for colour categorization in areas V1, V2 and V4 of macaque monkey cortex. Behavioral Brain Research, 76, 51-70.