Received: 6 July 2022 | Revised: 11 April 2023 | Accepted: 19 April 2023

DOI: 10.1111/desc.13406

Developmental Science WILEY

RESEARCH ARTICLE

# Colour perception changes with basic colour word comprehension

**Samuel H. Forbes**[1] | **Kim Plunkett**[2]

[1]Durham University, Durham, UK

[2]University of Oxford, Oxford, UK

**Correspondence**
Samuel H. Forbes.
Email: samuel.forbes@durham.ac.uk

**Funding information**
Rhodes Scholarships

## Abstract

Recent work has investigated the origin of infant colour categories, showing pre-linguistic infants categorise colour even in the absence of colour words. These infant categories are similar but not identical to adult categories, giving rise to an important question about how infant colour perception changes with the learning of colour words. Here we present two novel paradigms in which 12- and 19-month-old participants learning English as their first language were assessed on their perception of colour, while data on their colour word comprehension were also collected. Results indicate that participants' perception of colours close to the colour category boundaries dramatically change after colour word learning. The results highlight the shift made from infant colour categories to adult-like linguistically mediated colour categories that accompanies colour word learning.

KEYWORDS

categorical perception, colour words, eye-tracking, pupillometry

### Research Highlights

- We aimed to test whether colour perception is linguistically mediated in infants.
- We used novel eye-tracking and pupillometry paradigms to test infant colour perception either side of learning colour words.
- Infants' discrimination of colour changes after learning colour words, suggesting a shift due to colour word learning.
- A shift from pre-linguistic colour representation to linguistically mediated colour representation is discussed.

## 1 | INTRODUCTION

Colour perception develops extremely early in life. The basic psychophysical ability to distinguish colours has been shown to be in place by around 3 months of age (e.g. Morrone et al., 1993; Peeples & Teller, 1970, 1978; Teller, 1979, 1998). However, adults have been shown to perceive colour categorically (Bornstein & Korda, 1984; Daoutis et al., 2006; Kay & Kempton, 1984; Pilling et al., 2003; Roberson et al., 2000; Roberson & Davidoff, 2000), that is to say that they differentiate colours faster or more accurately when they belong to different categories than when they belong to the same category of colour, or are within a category boundary. But to what extent is our perception of colour affected by our knowledge of the colour vocabulary? We address these questions with infants who are yet to learn colour words, or have just started learning them to investigate discrepancies between the two groups, and determine whether language impacts colour perception.

In a later study (Gilbert et al., 2006), the role of language in perceptual discrimination of colour was investigated across the left and right visual fields, with the idea that the right visual field would be more influenced by language representations than the left. Gilbert et al. demonstrated with a reaction time task that adults show categorical perception (CP) of colour in the right visual field (RVF), equating to the left hemisphere of the brain. The fact that CP was found only in the hemisphere of the brain that mediated language use provided persuasive evidence that CP of colour was linguistically mediated, at least initially. The results of the study brought to light the possibility of powerful Whorfian influences of language on perception (for further discussion, see Regier & Kay, 2009; Regier & Xu, 2017).

Despite a large number of replications of the work with varying populations (Al-Rasheed et al., 2014; Al-Rasheed, 2015a, 2015b, 2016; Clifford et al., 2009; Drivonikou et al., 2007; Franklin et al., 2010; Gilbert et al., 2008; Goldstone & Hendrickson, 2009; Zhou et al., 2010), there have also been some mixed results in replicating the findings of Gilbert et al. on hemispheric effects of language-mediated colour perception (Brown et al., 2009; Jraissati, 2012; Lindsey & Brown, 2009; Ocelak, 2016; Witzel & Gegenfurtner, 2013, 2016). Some of these differences in findings might be the result of differences in colour spaces, or careful control of visual properties of the stimuli (see e.g. Witzel, 2016), or the use of the split visual field technique. Even without examining hemispheric effects, cross-linguistic research indicates that the presence of a category boundary may be dependent upon the size and boundaries of the colour *word* category in the native language of the speaker (e.g. Roberson et al., 2008, 2009). Due to the debate around hemispheric effects, one aim of the present study is to examine the role of learning colour words on colour perception, independent of hemispheric effects.

## 1.1 | Colour words and colour perception in infancy

While there is considerable debate around the hemispheric effects of language on CP of colour in adults, work in pre-linguistic infants provides a different perspective on the role of language in the development of colour categories. Categorical responding to colour stimuli has been consistently demonstrated in infants across a range of paradigms. In a series of experiments with English and Himba[1] toddlers, it was established that toddlers also display CP of colour, in much the same way as adults (Franklin et al., 2005a). This same CP of colour was also demonstrated with infants as young as 4 months of age (Franklin et al., 2005b), and using neuroimaging (e.g. Clifford et al., 2009; Yang et al., 2016) while similar effects were found with children of different language groups, up to around 7 years of age (Daoutis et al., 2006).

Although CP of colour is well-established in infancy, it is worth asking how well a participant needs to comprehend a colour word in order for it to change perception of colours close to the category boundary. Goldstein et al. (2009) demonstrated that only toddlers with a fuller knowledge of the colour categories showed categorical responding, both for Himba and English toddlers. The primary difference with the above work was that Goldstein et al. employed a much stricter definition of colour word knowledge, suggesting the level of familiarity with colour words may move those perceptual boundaries.

An important breakthrough in understanding the relationship between infant colour categories and adult colour categories was provided by Skelton et al. (2017), who demonstrated that infant colour categories are grounded in the biology of colour vision, while at the same time highlighting the similarities between infant colour categories and adult colour categories across various languages (although for further discussion on the links between infant and adult categories, see Siuda-Krzywicka et al., 2019). This indicates that there are likely similarities in the mechanisms that underlie infant colour categories and adult colour categories, without ruling out a role for language in colour perception. Indeed evidence from colour word learning studies indicates support for frequency of input and strength of comprehension of a colour word to be the catalyst for a shift in category boundary (Saji et al., 2020; Wagner et al., 2013, 2018; Yurovsky et al., 2015).

Looking at hemispheric effects again provides further insight into the influence of colour words on colour perception. In an extension of the hemispheric effects shown by Gilbert et al. (2006), similar effects were demonstrated with pre-linguistic infants. Using an eye-tracking paradigm, Franklin et al. (2008b) demonstrated that infants have a faster latency to the first look for the between-category colour. However, unlike the findings of Gilbert et al., infants in the study by Franklin et al. (2008b) demonstrated CP of colour in the left visual field, rather than the right, reversing the hemispheric lateralisation seen in adults. Despite the importance of the results and relatively large effect sizes, it is important to note that the results were based on only 26 pre-linguistic infants, half of whom were removed for fussiness.

A second study with toddlers, by the same authors (Franklin et al., 2008a), tested a larger sample of around 40 months of age. In that study, they were split into two groups based on their production of the colour terms (learners and namers). In concordance with the above study, the hemispheric lateralisation of CP was found to reverse with the learning of colour terms, analogous to the results found in infants versus adults in the previous study. Thus it appeared evident that pre-linguistic infants and toddlers possess colour categories and perceive colour categorically, and that these are related to but not identical to those of adults, exchanging early, biological colour categories for linguistically mediated CP with the learning of colour words (for similar results in memory see also Roberson et al., 2004).

In contrast to this result, however, by 40 months of age (or even within the full range of 2–5 years), toddlers would be expected to have a strong understanding of at least typical examples of the basic colour categories, suggesting that this distinction may reflect a more advanced comprehension or production, not their earliest comprehension (Forbes & Plunkett, 2019a, 2020; Wagner et al., 2018). Therefore, how changes in colour word comprehension – at its most basic level – influence perception of colour boundaries is worthy of further investigation (see also Forbes & Plunkett, 2019b, for evidence of very basic colour word comprehension influencing attention to coloured objects).

In Experiment 1, we examine how colour words might affect discrimination of colours straddling the colour category boundary in infants and toddlers, with a carefully controlled eye-tracking study that uses a forced-choice preferential looking task. In comparison to some previous work, a relatively large sample size (and lower drop-out rate due to the nature of the task) was tested to reinforce previous claims. Importantly, where this study differs from previous work is that we use parental report to capture early colour word comprehension (Forbes & Plunkett, 2020), one that has been cross-validated with eye-tracking in a large sample (Forbes & Plunkett, 2019a). In addition, to examine the role that colour words play in perceptual changes of colour, two age groups are tested: 12-month-olds, who are unlikely to know colour words based on prior work (Forbes & Plunkett, 2020), and 19-month-olds, for whom the colour word learning process is under way.

In Experiment 2, an innovative adaptation of previous infant looking time procedures is utilised. This is a paradigm commonly used during familiarisation in many categorisation studies, where infants are presented with one or more objects on a screen, and the total fixation time is used as a measure (e.g. Althaus & Plunkett, 2015a, 2015b; Plunkett et al., 2008; Younger, 1985). This is a robust paradigm, where total attention is measured, analogous to that also used in examining infant speech perception such as the Headturn Preference Procedure (Fernald, 1985; Johnson & Zamuner, 2010; Jusczyk & Aslin, 1995). In this study, the paradigm used is similar to the looking time mechanism used in both categorisation trials and in the Headturn Preference Procedure, but a single dynamic visual stimulus is presented with a constant auditory stimulus to help maintain attention, and the measure is the length of time infants attend to the visual stimulus as a result of the dynamic changes. In contrast to the Headturn Preference Procedure, where the image is kept constant but the sound changes, here the sounds are kept constant but the image changes. The dynamic stimulus of Experiment 2 changes between two colours – either within one colour category, or across two different colour categories. Crucially, the colour changes within any given dynamic stimulus are close to each other in colour space, and equal in luminance, which under carefully controlled experimental conditions allow the use of pupillometry, as well as the aforementioned looking time measurement. Pupillometry – the measure of changes in pupil size – while responsive to changes in light, also reflects attention and cognitive load, and can be reliably used with infants (Gredebäck & Melinder, 2011; Sirois & Jackson, 2012).

The two questions of interest, therefore, in Experiment 2 are whether we observe looking time differences when the visual stimulus changes between-category or within-category and whether that shifts when colour words are known, and whether we observe systematic pupil size changes across the same manipulations, and again, whether the patterns of pupil dilation seen change either side of learning colour words. Either of these changes would indicate that the perceptual response to the colour category boundary changes with the beginning of the process of learning colour words.

Note that across both experiments, we use Munsell stimuli, similar to many previous studies in the field (e.g. Franklin et al., 2008a, 2008b). Munsell stimuli were traditionally used to equate the stimuli between and within categories. More recent work has highlighted that Munsell may not be the most appropriate metric to determine category effects, and have instead used Just Noticeable Differences (e.g. He et al., 2014), or have used carefully controlled studies to show that category effects are a result of categorical status and not perceptual similarity (e.g. Skelton et al., 2017). Due to the choice of stimuli, differences across conditions cannot be interpreted as categorical, and so these experiments do not test CP. Rather, they are designed to test the role of language in the perceptual representation of colour, which can still clearly be seen as a change across conditions as a result of language learning with the stimuli used.

## 2 | EXPERIMENT 1

### 2.1 | Methods

#### 2.1.1 | Participants

A total of $N = 64$ participants all learning English as their native language took part in this study. Participants were recruited to the participant database either from the local maternity ward or online, and contacted to volunteer for the study. All participants lived in the Oxfordshire region; income and ethnicity were not recorded for this study. Participants were recruited for the study in two groups: 12-month-olds ($N = 33$, mean age 12.23 months, S.D. age 0.39 months), and 19-month-olds ($N = 31$, mean age 19.29 months, S.D. age 0.47 months). Of those, one 19-month-old participant was excluded due to failure to calibrate. The final analysed sample of participants was $N = 63$. No participant in this study recorded a family history of colour vision problems.

#### 2.1.2 | Materials

Stimuli for this experiment were carefully chosen, based on the analyses of Witzel and Gegenfurtner (2011, Study 1). In their re-analysis of Gilbert et al. (2006), the authors cite the need to keep both *Munsell value* and *chroma* constant for the stimuli, settling on a *Munsell value* of 5 and a *chroma* of 6 Fairchild (1998); Munsell Color Services (2007). The present study uses the same values, which are separated by five steps, described in Table 1. Four simulated *Munsell chips* were used (7.5G5/6, 2.5BG5/6, 7.5BG5/6, 2.5B5/6), using the CIE 1931 Yxy values given in Witzel and Gegenfurtner (2011, supplementary material). These were converted into a calibrated RGB for display on the monitor. These were the stimuli used in the *blue* and *green* conditions of this study.

Due to the concerns raised by Franklin et al. (2008b) that younger infants may not be responsive to fine differences in stimuli, an additional condition was included in this study. Stimuli for the *far* condition were instead separated by 10 steps, and designed based on the stimuli used in Franklin et al. (2008b). On this occasion, a constant *Munsell value* of 6 and a *chroma* of 8 were used, as in the aforementioned study. Illuminant C was created using the *x* and *y* values described by Fairchild (1998).

**TABLE 1** Chromaticity coordinates (CIE 1931) for the three conditions in the present study.

| Stimulus | Type | Y | x | y |
|---|---|---|---|---|
| 2.5B5/6 | Blue | 22.0 | 0.218 | 0.276 |
| 7.5BG5/6 | Blueish | 23.8 | 0.229 | 0.305 |
| 2.5BG5/6 | Greenish | 23.6 | 0.241 | 0.343 |
| 7.5G5/6 | Green | 23.1 | 0.260 | 0.368 |
| 2B6/8 | Blue (far) | 19.47 | 0.209 | 0.282 |
| 2BG6/8 | Greenish (far) | 19.47 | 0.235 | 0.355 |
| 2G6/8 | Green (far) | 19.47 | 0.286 | 0.431 |
| Illuminant C | | 23.8 | 0.310 | 0.316 |

In the *blue* condition, the first three are used; in the *green* condition, the second to fourth stimuli are used. The three denoted with 'far' are used in the *far* condition.

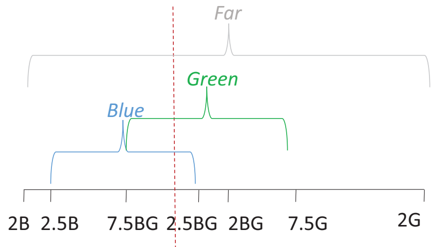


**FIGURE 1** Stimuli selections for each condition. The dashed line indicates an approximate category boundary.

### 2.1.3 | Design

Each condition was designed so that participants saw a background stimulus that was close to the category boundary between green and blue. Each background colour had two squares on it, each an equal number of steps from the background colour. One of those squares would be within the same colour category (within-category), and the other would cross the category boundary (between-category):

1. In the *blue* condition, participants saw a blueish (Table 1) background, with a greenish square on one side and a blue square on the other.
2. In the *green* condition, participants saw a greenish background, with a green square on one side and a blueish square on the other.
3. In the *far* condition, participants saw a greenish background (denoted with *far* in Table 1), with a green square on one side and a blue square on the other.

See Figure 1 for details of the colour samples chosen in each condition.

Each condition had four trials, with each colour appearing on the left twice and the right twice. This gave a total of 12 trials, which were randomised in order. To avoid participants habituating to any one colour, the screen was rendered Illuminant C between each trial for a minimum of 2 s. The monitor (1920 × 1080 pixels) used for presentation was calibrated using a datacolor Spyder 5 Elite calibration device, ensuring that the colours presented matched the intended hues. Each square was presented as a 531 × 531 pixel square, meaning that they each subsumed approximately 11° of visual angle.

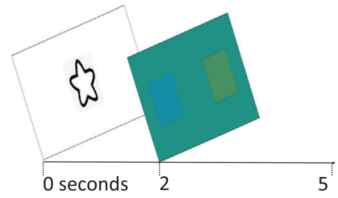


**FIGURE 2** A sample trial for Experiment 1. Borders around each square are for ease of viewing in this figure, and were not included in the experiment.

Given infants demonstrate CP of colour, as suggested in the literature (Franklin et al., 2005a, 2008a, 2008b), participants would be expected to show a preference for the between-category colour, although how this will change either side of learning the relevant colour words is unknown. To ensure that the findings are the result of colour word learning and not of experience or age, the effect of age will also be tested and compared with the effect of colour word knowledge.

### 2.1.4 | Procedure

After obtaining informed consent from the parents, and filling out the colour word supplement to the Oxford CDI (Forbes & Plunkett, 2020; Hamilton et al., 2000), and a short play session, participants were seated on the caregivers lap, approximately 75 cm from the screen. The standard 9-point calibration sequence was adapted to avoid priming the participant, using a solid black star moving across all 9 points, on a background approximating Illuminant C. Calibration was run until the eye-tracker successfully calibrated at least 7 of the 9 points.

At the beginning of each trial, a black and white flapping star oriented participant attention to the centre of the screen, which was triggered when the experimenter judged the infant to be looking at the screen. During that time, a recorded voice said '*Look!*'. After 2 s, the trial began, and the background colour and both squares appeared immediately on the screen. The trial continued for another 3 s. Figure 2 demonstrated the timeline of a typical trial. Infant gaze was measured with a Tobii TX300 remote eye-tracker, recording at 120 Hz. Trials were automated using a custom Matlab script.

### 2.1.5 | Analysis

Analysis was performed by first expanding areas of interest (the coloured squares – AOIs) by 25% to allow for noisier gaze data in infant tracking. Looking was calculated in each 100-ms time bin for each participant using the R package *eyetrackingR* (Forbes et al., 2021), focusing on the number of looks to each AOI. In any given timebin, then, we have calculated the number of looks to each AOI (and by extension the proportion of looking to one AOI over the other). These data were

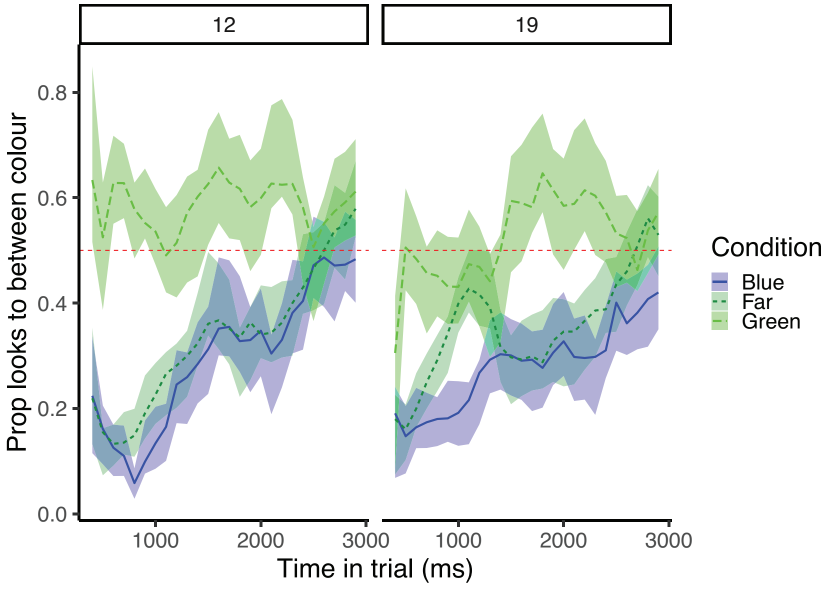


**FIGURE 3** Proportion looking to the between-category colour during the trial, split by age in months (12 left and 19 right). Coloured ribbons indicate the mean and standard error of raw data, while the coloured lines indicate model predicted mean. The red dashed line indicates chance looking at 0.5.

then analysed in such a way that we assess looking to the between-category AOI (the one that crosses the category boundary, so green on a blueish background or blue on a greenish background) as opposed to the within-category AOI, such that: $Prop_{between} = \frac{Looks_{between}}{Looks_{between}+Looks_{within}}$. Trials were removed where participants failed to contribute at least 10% of a trial's worth of looking data. This very lenient rule was set to avoid differential effects in each age group, with 12-month-old participants much more likely to lose trials at a more conservative threshold. A total of 170 trials, or just over 20% of the total, were removed due to this rule. Participants on average completed 9.24 trials (SD 2.18). The modelling of the data was done using the R package *glmmTMB* (Brooks et al., 2017).

## 2.2 | Results

### 2.2.1 | Condition and age group

Overall patterns of looking to the between-category colour can be demonstrated by the ribbons in Figure 3. In the *Blue* and *Far* (represented in blue and grey, respectively) conditions, there is a pattern of systematic looking to the within-category colour (i.e. away from the between-category colour). In the *Green* condition, however, there is perhaps some evidence of a systematic pattern of looking to the between-category colour which crosses the colour category threshold.

In order to assess the looking patterns in the data, we modelled the time course of the data with a generalised linear mixed-effect model. Because the data are proportion data, we selected a binomial model, to accurately capture the fact that the data can neither go below 0 or above 1. The dependent variable was the proportion of looks to the between-category object (calculated as described above), while the age group (difference coded) and the trial condition (*Blue*, *Far*, or *Green*) were included as predictors. In order to capture the shape of the looking patterns, we also included quartic polynomials of time in the trial, that is, time, time-squared, time-cubed and time to the power four. In order to avoid these four time terms auto-correlating with each other, they were orthogonal, or independently scaled and centred. The main predictors were allowed to interact with each other and each of the orthogonal time terms.

To allow for the different patterns of looking that each participant might exhibit, the model was fitted with a random intercept for the interaction between each participant and condition. Additionally, each participant and condition was also allowed a random slope for each of the four time terms, to allow for maximum flexibility in modelling the data.

The model output can be seen in Table 2. Of particular interest, the model indicates that there is strong evidence of looking differences as a result of condition, $\chi^2(2) = 49.127$, $p < 0.001$. In addition, Table 2 indicates that there is also strong evidence for different time course in each condition, evidenced by the strong interactions between each of the time terms and condition. However, there is no strong evidence for an effect of age on amount of looking to the between-target items, nor on the looking patterns, although there is a three-way interaction between linear time, age and condition. The model fit can be seen with the solid lines in Figure 3.

### 2.2.2 | Colour term knowledge

In addition to testing the differences by age and condition, we were particularly interested in whether colour word knowledge has an effect on the looking patterns of participants. We divided the participants into two groups, based on their knowledge of the relevant colour terms – *blue* and *green* – as reported by their caregiver (Forbes & Plunkett, 2020), rather than their age, as in the previous model. As we wanted to be able to assess the earliest form of colour word knowledge, we based it on their reported comprehension of either of those two terms based on the Oxford CDI. A participant who was reported to

**TABLE 2** Model terms assessing effect of condition and age on looking to the between-category colour.

| Variable | Chisq | DF | Pr(>Chisq) |
|---|---|---|---|
| (Intercept) | 33.208 | 1 | <0.001 |
| ot1 | 38.175 | 1 | <0.001 |
| ot2 | 7.707 | 1 | 0.006 |
| ot3 | 9.548 | 1 | 0.002 |
| ot4 | 1.480 | 1 | 0.224 |
| Age | 0.573 | 1 | 0.449 |
| Condition | 49.127 | 2 | <0.001 |
| ot1:Age | 0.518 | 1 | 0.472 |
| ot2:Age | 0.280 | 1 | 0.597 |
| ot3:Age | 0.414 | 1 | 0.520 |
| ot4:Age | 0.181 | 1 | 0.671 |
| ot1:Condition | 17.011 | 2 | <0.001 |
| ot2:Condition | 13.154 | 2 | 0.001 |
| ot3:Condition | 7.758 | 2 | 0.021 |
| ot4:Condition | 7.690 | 2 | 0.021 |
| Age:Condition | 4.961 | 2 | 0.084 |
| ot1:Age:Condition | 11.847 | 2 | 0.003 |
| ot2:Age:Condition | 1.039 | 2 | 0.595 |
| ot3:Age:Condition | 3.724 | 2 | 0.155 |
| ot4:Age:Condition | 2.0885 | 2 | 0.352 |

ot refers to each of the orthogonal time terms, such that ot1 is linear time, ot2 quadratic and so on. Effects tested with a type 3 Wald Chi-squared test.

**TABLE 3** Model coefficients of the binomial model testing the effect of colour term knowledge on looking.

| Variable | Chisq | DF | Pr(>Chisq) |
|---|---|---|---|
| (Intercept) | 23.526 | 1 | <0.001 |
| ot1 | 23.723 | 1 | <0.001 |
| ot2 | 8.315 | 1 | 0.004 |
| ot3 | 6.292 | 1 | 0.012 |
| ot4 | 1.035 | 1 | 0.309 |
| Know | 0.027 | 1 | 0.870 |
| Condition | 21.749 | 2 | <0.001 |
| ot1:Know | 0.010 | 1 | 0.922 |
| ot2:Know | 0.353 | 1 | 0.552 |
| ot3:Know | 0.011 | 1 | 0.918 |
| ot4:Know | 0.129 | 1 | 0.719 |
| ot1:Condition | 9.039 | 2 | 0.011 |
| ot2:Condition | 1.244 | 2 | 0.537 |
| ot3:Condition | 5.736 | 2 | 0.057 |
| ot4:Condition | 0.433 | 2 | 0.805 |
| Know:Condition | 7.070 | 2 | 0.029 |
| ot1:Know:Condition | 0.228 | 2 | 0.892 |
| ot2:Know:Condition | 17.671 | 2 | <0.001 |
| ot3:Know:Condition | 0.295 | 2 | 0.863 |
| ot4:Know:Condition | 7.906 | 2 | 0.019 |

Results show Wald's Chi-Squared on fixed effects. ot refers to each of the orthogonal time terms, such that ot1 is linear time, ot2 quadratic and so on.

comprehend either *blue* or *green* was classed as known, and one who knew neither term was classed as unknown. This meant that 15 participants were classed as knowing the colour words, and 47 participants as not knowing the colour word.

The data were modelled with a similar model to the previous one, with the exception that instead of including the participants' age group, this model included their colour term knowledge (difference coded), and interactions with condition and the time terms. Age was not included in this model due to the correlation between age and colour term knowledge, with only four 12-month-olds reporting knowledge of *blue* or *green*. One 12-month-old participant was removed from this analysis due to an incomplete Oxford CDI.

Model results can be seen in Table 3. Of particular interest to the question concern, the main effects and interactions including Knowledge (Know in Table 3). There was no evidence of a main effect of colour term knowledge, $\chi^2(1) = 0.027$, $p = 0.870$, however there was evidence of an interaction between colour term knowledge and condition $\chi^2(2) = 7.070$, $p = 0.029$. There were also two 3-way interaction between knowledge, condition and the quadratic and quartic time terms.

The patterns of looking to the between-category colour can be seen in Figure 4. There are small visible differences between participants who know the colour terms and those who do not know the colour terms in both the *far* condition, with a preference for the within-category colour, particularly in the first 2 s of the trial. In the *green* condition, and also the *blue* condition to a certain extent, we can see a definite pattern of systematic looking to the between-category colour for participants who know the colour terms from around the 1000–2500-ms time window, which is far more pronounced than in the participants who do not know either colour term. While the width of the standard error confirms the variability given the comparatively smaller number of participants, there is some evidence that colour word knowledge affects looking patterns in a colour perception task, at least in the *green* condition.

## 2.3 | Discussion

Experiment 1 highlights two main findings. First, we see that there are systematic patterns of looking, albeit an inconsistent one with our expectation of a preference for the between-category stimulus – a preference for the within-category colour in the *blue* and *far* conditions, and a preference for the between-category colour in the *green* condition. Second, we see that there is a possible effect of colour term knowledge on these looking patterns, with participants in one condition at least showing much more looking to the between-category colour.

These findings indicate that colour term knowledge might change the way an infant perceives the boundary of a colour category, with looking to the between-category colour much more likely when at least

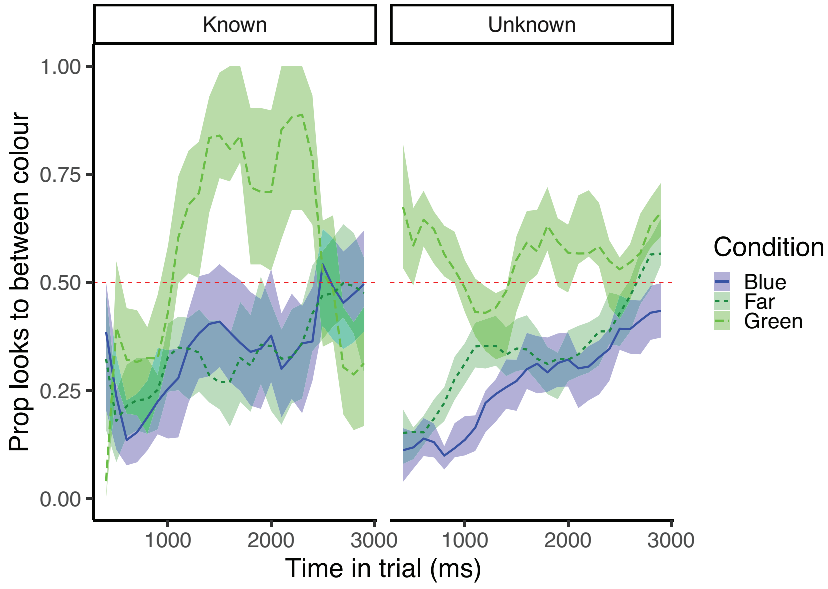


**FIGURE 4** Proportion looking to the between-category colour during the trial, split by colour term knowledge. Ribbons indicate the mean and standard error of raw data, while the lines indicate model predicted mean. The red dashed line indicated chance looking at 0.5.

one of the colour terms was marked as comprehended by the parents. While this finding was most obviously demonstrated in the *Green* condition, Figure 4 also suggests possible reduction in the preference to the within-category colour in the *Blue* condition as well.

Despite the interesting nature of these findings, an alternative explanation involving colour preference cannot be entirely ruled out. Forbes and Plunkett (2019a) found that infants and children across a range of age groups demonstrate a preference for green when contrasted with blue, although other studies reported longer looking overall to blue (Brown & Lindsey, 2013; Franklin et al., 2008c). Overall, the participants in Experiment 1 systematically looked to the blue object when presented against a green object on a green background (*Green* condition) and the blue object on a blue background when presented with a green object (*Blue* condition) – this would fit the explanation of a preference for blue. However, only in the *Far* condition where participants demonstrate preferential looking to the green object, making the colour preference explanation unlikely (although it does not rule out preferences for specific hues or saturations completely, see Brown & Lindsey, 2013, for discussion of this).

Experiment 1 indicates that infants do notice a colour crossing the colour category boundary, and that their perceptual space shifts with the learning of relevant colour names, but this finding is limited to static colour patches. For convergent evidence, Experiment 2 assesses whether they notice dynamic changes across a category boundary as opposed to those that remain in the same colour category.

## 3 | EXPERIMENT 2

### 3.1 | Methods

#### 3.1.1 | Participants

Participants in this experiment were $N = 62$ infants, 32 aged 12 months and 30 aged 19 months. All participants in this study also took part in the previous study, described above, except one participant who failed to start the task due to fussiness. There were an additional two participants who completed Experiment 1 but were unwilling to participate in this experiment due to fussiness.

Of those 62 total participants, $N = 4$ were removed having failed to start the task at all. Thus, the final analysable sample was $N =$ 30 19-month-olds (mean age 19.27 months, S.D. age 0.47 months), and $N = 28$ 12-month-olds (mean age 12.28 months, S.D. age = 0.37 months).

**TABLE 4** Chromaticity coordinates (CIE 1931) for the three conditions in the present study.

| Stimulus | Type | *Y* | *x* | *y* |
|---|---|---|---|---|
| 2.5B5/6 | Blue | 22.0 | 0.218 | 0.276 |
| 7.5BG5/6 | Blue-green | 23.8 | 0.229 | 0.305 |
| 2.5BG5/6 | Green-blue | 23.6 | 0.241 | 0.343 |
| 7.5G5/6 | Green | 23.1 | 0.260 | 0.368 |
| Illuminant C | | 23.8 | 0.310 | 0.316 |

#### 3.1.2 | Materials

The same Munsell tiles were used for this experiment. Four *Munsell tiles* (Munsell Color Services, 2007) were simulated using CIE information provided in Witzel and Gegenfurtner (2011, supplementary material). These stimuli were created with a *Munsell value* of 5 and a *chroma* of 6 (Fairchild, 1998), and each stimulus was separated by five steps. There were two blue squares, one of which (Blue-green) was close to the category boundary, and two green squares, with one (Green-blue) close to the category boundary. Additionally, Illuminant C was created using the *x* and *y* values described by Fairchild (1998). The stimuli are described in Table 4.

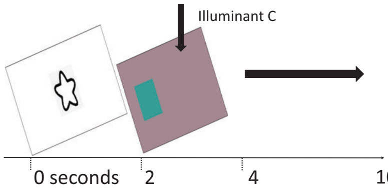


**FIGURE 5** A typical trial for Experiment 2. The patch of colour would alternate between two shades until the trial ends, every 500 ms.

### 3.1.3 | Design

The four stimuli listed above were combined to make three conditions.

- The *within-blue* condition alternated between presenting the Blue stimulus and the Blue-green stimulus.
- The *within-green* condition alternated between presenting the Green stimulus and the Green-blue stimulus.
- The *between* condition alternated between presenting the Blue-green stimulus and the Green-blue stimulus.

Each stimulus pair was shown alternating every 500 ms for 8 s, accompanied by an auditory stimulus consisting of three tones presented repeatedly in ascending order on loop throughout the trial, designed to maintain the infants' attention to the task. The same auditory stimuli were presented for all conditions, with the ascending tones not designed to be synchronous to the changing colours.

### 3.1.4 | Procedure

The procedure of this experiment took place alongside Experiment 1, counterbalanced to avoid fatigue becoming a major factor in the results. At the beginning of each trial, an attention-getter was presented in the middle of the screen for 2000 ms, to bring fixation back to the middle of the screen. The trial began immediately after, with the trial images appearing on the screen immediately. Stimuli were presented as a square (531 × 531 pixels) on the screen, which alternated between the two colours listed for each every 500 ms. In order to avoid infants looking purely because it was in their gaze, stimuli were presented on one side of the screen, rather than in the middle. The background of the trial was kept to Illuminant C in order to standardize and limit habituation to any of the colours in the trial (Fairchild, 1998), and following the 8 s of trial, Illuminant C was shown for a minimum of 2 s before the start of the next trial (see Figure 5). The monitor (1920 × 1080 pixels) used for presentation was calibrated using a datacolor Spyder 5 Elite calibration device.

Each participant saw 12 trials. In order to maintain maximal attention throughout the experiment, six trials were presented on one side of the screen, and then six trials were presented on the other side of the screen. Each block of six trials contained two trials from each condition, counterbalanced by starting stimulus. Trials within each block were presented in a random order. Infant looking data were collected by a Tobii TX300 eye-tracker operating at 120 Hz, and trials were automated with a custom MATLAB script.

### 3.1.5 | Analysis

Methods of analysis for infant looking time experiments differs greatly across experiments and design, and thus great care was taken in analysing this experiment. Data were first extracted with custom scripts in MATLAB, which passed the infant looking data through a fixation filter, where a fixation was defined as looking in the same location (with a certain amount of dispersion to allow for the instability of infant gaze) for a minimum of 100 ms (for a discussion of fixation filters, and why they are debated in looking time analyses, see Wass et al., 2013).

As total looking time in each trial is the primary measurement of this experiment, as opposed to an indicator of category learning found in habituation trials of categorisation experiments, outliers were not removed as an a priori decision.[2] However, since it was still important to remove trials where participants had no interest at all in the task, trials with greater than 90% trackloss were removed from the analysis. The seven participants who were removed for failure to complete a trial, did not complete a trial in which trackloss was less than 90%. Gaze data were analysed using the *eyetrackingR* package (Forbes et al., 2021) and lme4 (Bates et al., 2015) in R.

A key measure of this study is also to measure the change in pupil size. As the luminance is kept constant even as the colour changes, and due to the carefully controlled nature of the task, this study is ideal for pupillometric analysis. Pupil size data were cleaned and filtered using the R package *PupillometryR* (Forbes, 2020). Specifically, the data were filtered with a Hanning filter of degree 11, downsampled to 50-ms time bins to reduce autocorrelation, and the first 1000 ms were removed to allow for the initial pupil constriction that occurs when the object appears on screen. The pupil size in the first 50 ms immediately thereafter was taken as a baseline which was subtracted from the pupil size in all subsequent time bins, such that a pupil size of 0 represents no change in pupil size from baseline. The final 2000 ms were also removed due to decreasing numbers of participants who contributed data during this window.

## 3.2 | Results

### 3.2.1 | Gaze data

We initially calculated the overall proportion of the trial the participants were willing to maintain attention to the object on screen. The overall proportion of trial time spent looking to the target can be visualised as a raincloud plot (Allen et al., 2019) in the top panel of Figure 6. Overall, there appear to be minimal differences between the trials that cross a colour category boundary (between: in pink) versus those that do not (within: in teal).

We modelled these data with a linear mixed-effects regression, where the proportion of time spent looking to the target was the

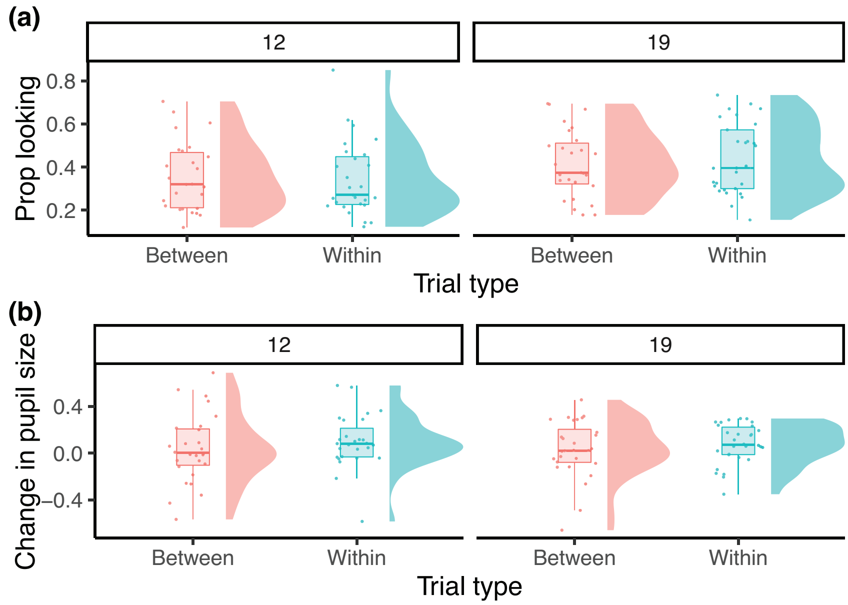


**FIGURE 6** (a) Overall looking to the target during Experiment 2. (b) Overall pupil size during Experiment 2. Both (a) and (b) are split by age in months (12 left and 19 right). Note the data are presented as a raincloud plot (Allen et al, 2019), where the density of the data appear on the right as a half-violin plot, the median and interquartile range appear on the left as a boxplot, and the raw data points are superimposed on top of the boxplot, with random left-right jitter.

**TABLE 5** Output of Wald's type 3 Chi-squared test on a linear model assessing the effects of age and trial type on time spent looking to the target.

| Variable | Chisq | DF | Pr(>Chisq) |
|---|---|---|---|
| (Intercept) | 388.67 | 1 | <0.001 |
| Type | 0.110 | 1 | 0.740 |
| Age | 4.145 | 1 | 0.042 |
| Type:Age | 0.604 | 1 | 0.437 |

dependent variable. The independent variables were the difference-coded trial type (within and between) and age cohort (12 or 19 months). We also included the interaction between the two independent variables. Due to the fact that each participant saw each trial type, we also included a random intercept for each participant (a random slope was not possible with the number of data points used). The results, shown in Table 5, indicated no strong evidence for a difference due to trial type, but some evidence for an effect of age, $\chi^2(1) = 4.145, p = 0.042$.

To assess the effects of colour word learning on overall looking to the target, one participant who did not have reported CDI data for their colour knowledge was removed and the model re-run. A second model was then run, adding the additional independent variable of colour term knowledge, as well as all interactions. This model did not improve model fit, indicating no effect of colour term knowledge or any interactions with the other variables, $\chi^2(4) = 2.206, p = 0.698$.

### 3.2.2 | Pupil data

Overall pupil size in each trial type and age group can be seen in the lower panel of Figure 6. Due to the richness and variability in the pupil size data over time, data analyses were completed on the time course data which indicates change in pupil size over time from baseline, rather than aggregated values. In order to assess whether there are differences in pupil size as a function of trial type (within or between-category), we used functional data analysis (Ramsay & Silverman, 1997). The average pupil size in each 50-ms timebin for each infant in each trial type was calculated (Figure 7a), and then the between-trial averages were subtracted from the within trial averages, leaving a time course of differences between conditions for each participant (Figure 7b). These curves were then fit with order 4 B-splines (for a full overview of these techniques see Jackson & Sirois, 2009; Sirois & Jackson, 2012) with 10 bases, chosen because it fitted the major characteristics of the data (Figure 7c). We then performed a functional *t*-test on the smoothed data, which allow us to compute a single *t*-test on an event as it unfolds over the time course of a trial. The results can be seen in Figure 7d, where the highlighted region indicates the time period where the *t* value crosses the critical threshold, indicating evidence of an effect of a difference in trial type at an alpha of 0.05 (around 1800–2900 ms), such that participants have a greater change in pupil size in the within-colour trials than the between-colour trials.

In order to account for the effects of colour term knowledge, we needed a complex modelling procedure that would allow for careful analysis of the interaction between trial type and colour term knowledge, while preserving the nonlinear effects that individual differences have on the pupil dynamics over the course of the trial, so we opted for a Generalised Additive Model.

We structured the model so that a separate factor smooth was used for each level of the interaction between participant ID and trial type, allowing us to control for the different pupil dynamics each individual might undergo depending on the trial type. We created a base model with only a fixed intercept, and due to the fact that the density of the pupil data suggests a fat-tailed distribution, used the scaled-*t* model family rather than the Gaussian family. We then extracted the starting temporal autocorrelation value to use in subsequent models to

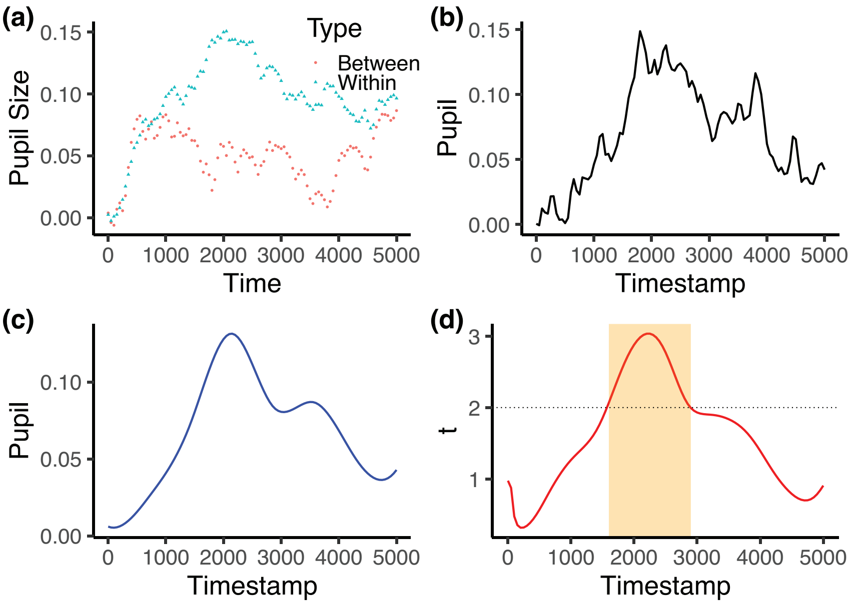


**FIGURE 7** The steps in functional data analysis of pupil data. (a) The change in pupil sizes over time for each trial type. (b) The difference curve of within–between. (c) The difference curve fitted with a smoothing spline. (d) The results of a functional *t*-test on the smoothed difference, where the highlighted region shows $p < 0.05$.

control for the fact that pupil size at time $t$ is correlated with pupil size at both time $t + 1$, and also time $t - 1$. We re-ran the base model with an AR1 model in the data based on the initial autocorrelation value (for detailed discussion of the importance of this with pupil data, see van Rij et al., 2019).

We then added in a separate smooth of colour word knowledge over time, which improved model fit as assessed by change in fREML (van Rij et al., 2020), $\chi^2(3) = 4.690$, $p = 0.025$ indicating evidence of an overall difference in looking due to colour word knowledge. We, subsequently, added in a smooth of trial type by time, which also improved model fit, $\chi^2(3) = 19.999$, $p < 0.001$. Finally, we added a parametric coefficient for the interaction (note that a parametric coefficient is not required for difference smooths, see van Rij et al., 2019) which indicated evidence for an interaction between knowledge and trial type $\chi^2(3) = 6.992$, $p = 0.003$. Adding an additional smoother to the interaction term did not improve model fit, $\chi^2(8) = -127.758$, and as such was not included in the final model. The final model fit was adequate ($R^2_{adj} = 0.94$).

The fit from the final GAMM model to the raw pupil data can be seen in Figure 8. As can be seen from the plot, participants who were classed as knowing either *green* or *blue* exhibited a greater change in pupil size from baseline in the between-category condition than the within-category condition. Conversely, those who knew neither colour term showed a substantial change in pupil dilation in the within-category condition than the between-category condition.

### 3.3 | Discussion

Experiment 2 again demonstrates two main findings using a novel paradigm to test for colour category representation. First, the pupil data indicate overall differences in the change in pupil size as a function of trial type. Specifically, on average, participants had a greater increase in pupil diameter when watching the coloured squares change between two colours within the same category (within-category trial type) than when watching the two squares change by the same amount in terms of Munsell steps, but across a category condition (the between-category trial type).

Second, further investigation into the pupil data using GAMMs indicated that the overall effect of trial type might depend on the participants own knowledge of colour terms. Participants marked as comprehending either *blue* or *green* – the relevant colour categories under examination – in fact demonstrated greater changes in pupil size when the colour squares crossed a colour category boundary. Further, the interaction effect highlights that this change in pupil dynamics over different trial types is a function of colour term knowledge, suggesting that even an early understanding of basic colour words as indicated in parental reports (Forbes & Plunkett, 2020, 2019a) affects the understanding of colour boundaries.

These findings seen in the pupil size dynamics were not in turn reflected in the gaze data, which is not altogether surprising. Total looking time averaged across trials is a rough measure, whereas by comparison, the time course of pupil size changes is a very sensitive measure (Sirois & Jackson, 2012), and more likely to capture changing dynamics in cognition.

## 4 | OVERALL DISCUSSION

In the present study, we devised two novel experimental paradigms to test infants' changing colour perception across the critical age threshold where they begin to learn colour terms, and whether that perception is altered by the colour words they know. In Experiment 1, a preferential looking paradigm was used to allow participants to choose to look at one of two coloured squares – both equidistant from the colour of the background in Munsell hue steps – where one of

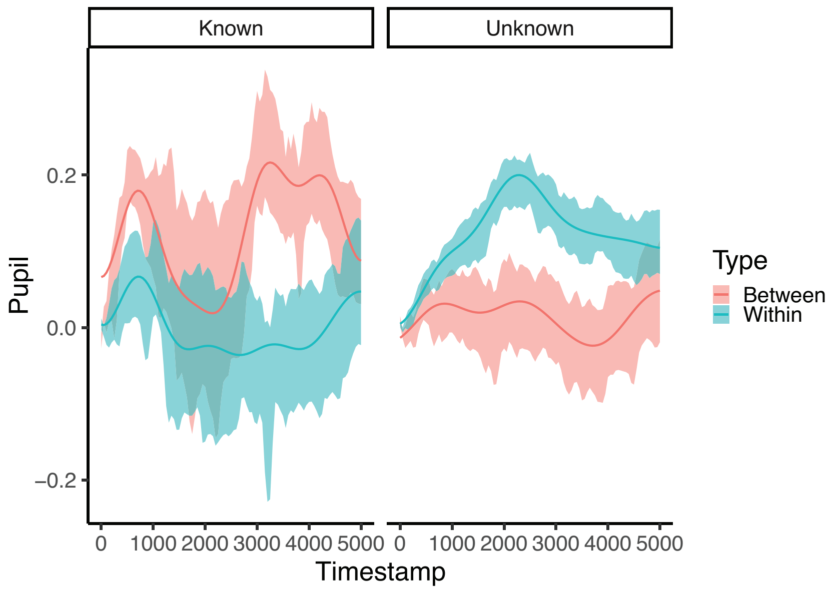


**FIGURE 8** Changes in pupil size over the course of the trial, split by trial type and colour word knowledge. Ribbons indicate mean and standard error of the raw data, solid lines indicate model fit from the model including knowledge and trial type, as well as the parametric interaction.

these squares crossed the colour boundary as typically understood by adults. The findings indicated systematic patterns of looking, but exhibiting different patterns across conditions, with a within-category preference in two of the three conditions. Additionally, learning colour words had an effect on looking in at least one condition, encouraging more looking to the between-category colour, suggesting a role for colour word knowledge in colour category recognition and perception.

In Experiment 2, participants saw a single block of colour which changed constantly between two hues over the course of the trial. The Munsell distance of the change was kept equal across both trial types, but the change could either occur between or across a colour category as typically understood by English-speaking adults. Tracking pupil size changes over the course of the trial indicated that while in general, participants were more prone to an increase in pupil size when looking at a trial that changed within-category, this trend reversed when they knew at least one of the colour terms involved, showing the greatest increase in pupil size when watching the block change across a colour boundary.

Taken together, the findings indicate a changing representation of colour in infants and toddlers before and after early comprehension of colour terms, that seemingly shifts *with* the acquisition of these colour terms, shifting towards a between-category effect with colour term knowledge. Recall that in Experiment 1, there was a consistent preference in two conditions for the within-category colour, but that looking to the between-category colour increased in some conditions when colour words were learned. Similarly, in Experiment 2, a clear difference could be seen in pupil size changes from baseline when looking at within- versus between-category colour changes, but this reversed when the colour words were known. Thus infants seem to perceive colour differentially with the acquisition of the relevant basic colour terms. This supports claims that infants move from a biological and perceptual infant colour representation (Franklin & Davies, 2004; Skelton et al., 2017) and slowly extend the border of their colour representation toward linguistically mediated adult-like representations of colour (Forbes & Plunkett, 2020; Wagner et al., 2013).

It is important to keep in mind that the two experiments yield different measurements, and therefore, by extension, a different understanding of the effect. In Experiment 1, we see that overall, participants switch from having a within-category preference (that is to say they prefer the colour in the same colour category) towards a between-category preference with the knowledge of colour words. In the *Blue* and *Far* conditions, participants still have a within-category preference, although this is less exaggerated when they know the colour word than when they do not. In the *Green* condition, a between-category preference becomes much more pronounced when the colour word is known, indicating an overall preference shift toward the between colour with the knowledge of colour words. In Experiment 2, we instead see that participants had a greater increase in pupil size – a measure of cognitive effort or arousal – when seeing within-category changes if they did not know the colour words, and in between-category changes if they did. Each of these results are consistent with a shifting recognition of the colour category boundary caused by changes in the perceptual space, and are convergent across these very different measures.

Previous findings, particularly when examining hemispheric effects of colour word learning on CP, have indicated CP may change alongside learning of colour words, where lateralization of the CP changed from the right hemisphere to the left hemisphere as colour words are learned (Franklin et al., 2008a, 2008b). Using entirely separate paradigms, without examining hemispheric effects, we find support for the same basic idea that the perceptual representation of colour changes as colour words are learned, using both gaze data and pupillometric measures. This also goes some way toward further highlighting the clear shift from infant, biological colour categories to adult-like, linguistically mediated colour categories. Further, it highlights that this shift to linguistically mediated representations comes at the earliest, most basic understanding of a colour word, not with an adult-like representation. It is important to note that the present study was not designed to follow up the question of hemispheric effects directly. The two paradigms as measured here follow gaze and pupil size over the course of the trial, but once the participant has centrally fixated a target, the object is then in the centre of their vision, meaning that hemispheric effects cannot be measured using the methods employed in this study.

A key idea in this debate is that of colour term comprehension. What does it mean to truly comprehend a colour word? In the present study, we take parental report of comprehending at least one colour term used in the study. Parental report provides a sensitive indication of early comprehension (Forbes & Plunkett, 2020) that is commensurate with comprehension measured in tightly controlled behavioural experiments (Forbes & Plunkett, 2019a). In our view, the important step to triggering the move to a linguistically mediated representation of colour is that the child begins to understand the very basics of the colour category centroid, not a full adult-like comprehension of the term (Wagner et al., 2018), and thus using this more sensitive measure to capture early comprehension is appropriate. This is also consistent with evidence that the colour category boundary may shift with greater frequency of hearing the colour word (Saji et al., 2020). Similar findings were noted in the domain of memory, where Roberson et al. (2004) found that the effect of naming colours on memory increased with age, indicating that same shift from perceptual colour categories to a more linguistically mediated set.

An additional finding in Experiment 2 is that the pupil data show evidence of a shift in perceptual colour space that is mediated by colour word comprehension, but the gaze data do not. Sirois and Jackson (2012) highlight the role that pupil data may play in providing more direct physiological evidence than raw looking times, suggesting a stronger measure. Beyond this claim, it is also true that data that retain all the rich time-course information will be a more powerful measure than an aggregate dataset. Time-course analyses of gaze data also provide a much richer interpretation of behaviour than aggregation (Barr, 2008; Mirman, 2014), as we also see in Experiment 1, so it comes as no surprise that the less sensitive measure does not detect the effect, and a time-course analysis of gaze data when there is only one target, as in Experiment 2, is unlikely to be informative.

One difficulty with measuring infant perception under such tightly controlled conditions is that there is the potential to habituate to a certain stimulus set, warping perceptual properties. In both experiments, we attempted to mitigate this by always displaying a colour on the screen close to Illuminant C for 2 s between trials, in order to keep the participants from habituating. We cannot be certain, however, that this fully mitigates warping over the course of the experiment. It is also important to note that in the present study, only the blue-green boundary is examined. Previous studies have investigated some of the other boundaries (e.g. Drivonikou et al., 2007), and further work should be undertaken to examine the nature of the shift to linguistically mediated colour categories across the full set of colours.

Further, this experiment used the Munsell colour space to ensure equidistant stimuli in Munsell hue steps, based on stimuli used in previous research in order to best represent those findings. Using Munsell approximates the uniform colour space, but it is an imperfect approximation (for discussion of this see Witzel, 2016). While the effect of change in perceptual space within stimulus that we observe across both experiments as a result of learning colour names is not affected by this, less can be gleaned about what this means for CP of colour. A stricter test of this would use Just Noticeable Differences (JNDs; He et al., 2014) to have greater control of the stimuli. This is one possible cause for the overall preference we see for the within-category effect, although this is also reflected in the fact that it is consistently participants without colour word knowledge that demonstrated a within-category effect, and these participants were greater in number, due to the fact that nearly all 12-month-olds did not know the colour words, but not all 19-month-olds did.

Finally, one cannot rule out societal factors as a mechanism here. Demographic data on the income or educational status of participants were not collected as part of this study, but it is possible that participants who knew colour words had a different socioeconomic status from those who did not. This may have resulted in earlier colour word learning and a larger vocabulary, or even in an extreme case, differing visual statistics of their environments. This seems an unlikely mechanism but one not ruled out entirely by the current design. Likewise, the results reported here are constrained by the participant sampling, where we were restricted to inviting children learning English in proximity to the lab in Oxfordshire. Further research should look to see whether the effect of naming we show here holds across languages, especially in light of the recent findings that demonstrate that cultural factors underlie the growth in number of colour terms of the Himba (Mylonas et al., 2022).

Across the two experiments presented in this study, we demonstrate strong and consistent evidence that infants and toddlers see a shift in their perceptual colour space with the acquisition of colour words. Evidence shown here with gaze and pupillometric data further highlight that it is the emergence of the very basic colour word knowledge that is the catalyst for this shift, rather than simply a factor of age or experience.

## ACKNOWLEDGMENTS

The work was funded by a Rhodes Scholarship awarded to SHF. The authors are grateful to Irina Lepadatu and the Oxford Babylab team for support and thoughtful discussion, and are indebted to participants at the LCICD 2022 conference for their feedback, as well as the helpful suggestions of two anonymous reviewers.

## CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

## DATA AVAILABILITY STATEMENT

Data and analysis code are available at: https://osf.io/y39xq/

## ORCID

*Samuel H. Forbes* https://orcid.org/0000-0003-1022-4676
*Kim Plunkett* https://orcid.org/0000-0003-0216-7480

## ENDNOTES

[1] An indigenous northern Namibian small-scale society, who also live in Angola. Himba was at the time reported to have only five basic colour terms at the time of testing, as opposed to the 11 existent in English (Franklin et al., 2009), although see Mylonas et al. (2022) for recent updates.

[2] Treatment of outliers vary greatly across this type of analysis, but if participants who look less than a certain amount are excluded, then there can be no variability in looking time. Removing outliers also appears to have no major impact on the results.

## REFERENCES


Al-Rasheed, A. S. (2015a). Categorical perception of color: Evidence from secondary category boundary. *Psychology Research and Behavior Management*, *8*, 273–285. https://doi.org/10.2147/PRBM.S78348

Al-Rasheed, A. S. (2015b). An experimental study of gender and cultural differences in hue preference. *Frontiers in Psychology*, *6*, 1–5. https://doi.org/10.3389/fpsyg.2015.00030

Al-Rasheed, A. S. (2016). Categorical perception of colour is lateralised to the left hemisphere: Evidence from present-absent and pop-out tasks. *International Journal of Social Science and Humanity*, *6*(2), 113–118. https://doi.org/10.7763/IJSSH.2016.V6.629

Al-Rasheed, A. S., Franklin, A., Drivonikou, G. V., & Davies, I. R. L. (2014). Left hemisphere lateralization of categorical color perception among Roman and Arabic script readers. *Psychology*, *5*, 255–270.

Allen, M., Poggiali, D., Whitaker, K., Marshall, T. R., & Kievit, R. A. (2019). Raincloud plots: A multi-platform tool for robust data visualization. *Wellcome Open Research*, *4*, 1–41. https://doi.org/10.12688/wellcomeopenres.15191.1

Althaus, N., & Plunkett, K. (2015a). Categorization in infancy: Labeling induces a persisting focus on commonalities. *Developmental Science*, 1–11. https://doi.org/10.1111/desc.12358

Althaus, N., & Plunkett, K. (2015b). Timing matters: The impact of label synchrony on infant categorisation. *Cognition*, *139*, 1–9. https://doi.org/10.1016/j.cognition.2015.02.004

Barr, D. J. (2008). Analyzing 'visual world' eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, *59*(4), 457–474. https://doi.org/10.1016/j.jml.2007.09.002

Bates, D. M., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

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*(3), 207–222. https://doi.org/10.1007/BF00308884

Brooks, M. E., Kristensen, K., van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., Skaug, H. J., Maechler, M., & Bolker, B. (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R Journal*, *9*(2), 378–400.

Brown, A. M., & Lindsey, D. T. (2013). Infant color vision and color preferences: A tribute to Davida Teller. *Visual Neuroscience*, *30*(5-6), 243–250. https://doi.org/10.1017/S0952523813000114

Brown, A. M., Lindsey, D. T., Rambeau, R. S., & Shamp, H. A. (2009). Visual search for colors as a test of the Sapir-Whorf hypothesis. *Journal of Vision*, *9*(8), 366.

Clifford, A., Franklin, A., Davies, I. R. L., & Holmes, A. (2009). Electrophysiological markers of categorical perception of color in 7-month old infants. *Brain and Cognition*, *71*(2), 165–172. https://doi.org/10.1016/j.bandc.2009.05.002

Daoutis, C. A., Franklin, A., Riddett, A., Clifford, A., & Davies, I. R. L. (2006). Categorical effects in children's colour search: A cross-linguistic comparison. *British Journal of Developmental Psychology*, 373–400. https://doi.org/10.1348/026151005X51266

Drivonikou, G. V., Kay, P., Regier, T., Ivry, R. B., Gilbert, A. L., Franklin, A., & Davies, I. R. L. (2007). Further evidence that Whorfian effects are stronger in the right visual field than the left. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(3), 1097–1102. https://doi.org/10.1073/pnas.0610132104

Fairchild, M. D. (1998). *Color appearance models*. Addison-Wesley.

Fernald, A. (1985). Four-month-old infants prefer to listen to motherese. *Infant Behavior and Development*, *8*(2), 181–195. https://doi.org/10.1016/S0163-6383(85)80005-9

Forbes, S. H. (2020). PupillometryR: An R package for preparing and analysing pupillometry data. *Journal of Open Source Software*, *5*(50), 2285. https://doi.org/10.21105/joss.02285

Forbes, S. H., Dink, J., & Ferguson, B. (2021). eyetrackingR. R package version 0.2.0. http://www.eyetracking-r.com/

Forbes, S. H., & Plunkett, K. (2019a). Infants show early comprehension of basic color words. *Developmental Psychology*, *55*(2), 240–249. https://doi.org/10.1037/dev0000609

Forbes, S. H., & Plunkett, K. (2019b). The role of colour labels in mediating toddler visual attention. *Cognition*, *186*, 159–170. https://doi.org/10.1016/j.cognition.2019.01.008

Forbes, S. H., & Plunkett, K. (2020). Linguistic and cultural variation in early color word learning. *Child Development*, *91*(1), 28–42. https://doi.org/10.1111/cdev.13164

Franklin, A., Catherwood, D., Alvarez, J., & Axelsson, E. (2010). Hemispheric asymmetries in categorical perception of orientation in infants and adults. *Neuropsychologia*, *48*(9), 2648–2657. https://doi.org/10.1016/j.neuropsychologia.2010.05.011

Franklin, A., Clifford, A., Williamson, E., & Davies, I. R. L. (2005a). Color term knowledge does not affect categorical perception of color in toddlers. *Journal of Experimental Child Psychology*, *90*(2), 114–141. https://doi.org/10.1016/j.jecp.2004.10.001

Franklin, A., & Davies, I. R. L. (2004). New evidence for infant colour categories. *British Journal of Developmental Psychology*, *22*(3), 349–377. https://doi.org/10.1348/0261510041552738

Franklin, A., Drivonikou, G. V., Bevis, L., Davies, I. R. L., Kay, P., & Regier, T. (2008a). Categorical perception of color is lateralized to the right hemisphere in infants, but to the left hemisphere in adults. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(9), 3221–3225. https://doi.org/10.1073/pnas.0712286105

Franklin, A., Drivonikou, G. V., Clifford, A., Kay, P., Regier, T., & Davies, I. R. L. (2008b). Lateralization of categorical perception of color changes with color term acquisition. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(47), 18221–18225. https://doi.org/10.1073/pnas.0809952105

Franklin, A., Pilling, M., & Davies, I. R. L. (2005b). The nature of infant color categorization: Evidence from eye movements on a target detection task. *Journal of Experimental Child Psychology*, *91*(3), 227–248. https://doi.org/10.1016/j.jecp.2005.03.003

Franklin, A., Pitchford, N., Hart, L., Davies, I. R. L., Clausse, S., & Jennings, S. (2008c). Salience of primary and secondary colours in infancy. *British Journal of Developmental Psychology*, *26*, 471–483. https://doi.org/10.1348/026151007X256672

Franklin, A., Wright, O., & Davies, I. R. L. (2009). What can we learn from toddlers about categorical perception of color? Comments on Goldstein, Davidoff, and Roberson. *Journal of Experimental Child Psychology*, *102*(2), 239–245. https://doi.org/10.1016/j.jecp.2008.08.003

Gilbert, A. L., Regier, T., Kay, P., & Ivry, R. B. (2006). Whorf hypothesis is supported in the right visual field but not the left. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(2), 489–494. https://doi.org/10.1073/pnas.0509868103

Gilbert, A. L., Regier, T., Kay, P., & Ivry, R. B. (2008). Support for lateralization of the Whorf effect beyond the realm of color discrimination. *Brain and Language*, *105*(2), 91–98. https://doi.org/10.1016/j.bandl.2007.06.001

Goldstein, J., Davidoff, J., & Roberson, D. (2009). Knowing color terms enhances recognition: Further evidence from English and Himba. *Journal of Experimental Child Psychology*, *102*(2), 219–238. https://doi.org/10.1016/j.jecp.2008.06.002

Goldstone, R. L., & Hendrickson, A. T. (2009). Categorical perception. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(1), 69–78. https://doi.org/10.1002/wcs.026

Gredebäck, G., & Melinder, A. (2011). Teleological reasoning in 4-month-old infants: Pupil dilations and contextual constraints. *PLoS ONE*, *6*(10), 2–6. https://doi.org/10.1371/journal.pone.0026487

Hamilton, A., Plunkett, K., & Schafer, G. (2000). Infant vocabulary development assessed with a British communicative development inventory. *Journal of Child Language*, *27*(3), 689–705. https://doi.org/10.1017/S0305000900004414

He, X., Witzel, C., Forder, L., Clifford, A., & Franklin, A. (2014). Color categories only affect post-perceptual processes when same- and different-category colors are equally discriminable. *Journal of the Optical Society of America. A, Optics, Image Science, and Vision*, *31*(4), 322–31. https://doi.org/10.1364/JOSAA.31.00A322

Jackson, I. R., & Sirois, S. (2009). Infant cognition: Going full factorial with pupil dilation. *Developmental Science*, *12*(4), 670–679. https://doi.org/10.1111/j.1467-7687.2008.00805.x

Johnson, E. K., & Zamuner, T. (2010). Using infant and toddler testing methods in language acquisition research. In E. Blom, & S. Unsworth (Eds.), *Experimental methods in language acquisition research* (pp. 73–93). John Benjamins Publishing Company.

Jraissati, Y. (2012). Categorical perception of color: Assessing the role of language. *Croatian Journal of Philosophy*, *XII*(36), 439–463.

Jusczyk, P. W., & Aslin, R. N. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology*, *29*(1), 1–23. https://doi.org/10.1006/cogp.1995.1010

Kay, P., & Kempton, W. (1984). What is the Sapir-Whorf hypothesis? *American Anthropologist*, *86*(1), 65–79. https://doi.org/10.1525/aa.1984.86.1.02a00050

Lindsey, D. T., & Brown, A. M. (2009). Color difference scaling at the blue-green color category boundary as a test of the Sapir-Whorf Hypothesis. *Journal of Vision*, *9*(8), 340.

Mirman, D. (2014). *Growth curve analysis and visualization using R analysis and visualization using R*. Chapman & Hall/CRC Press.

Morrone, M. C., Burr, D. C., & Fiorentini, A. (1993). Development of infant contrast sensitivity to chromatic stimuli. *Vision Research*, *33*(17), 2535–2552. https://doi.org/10.1016/0042-6989(93)90133-H

Munsell Color Services (2007). *The Munsell book of color: Glossy collection*. X-rite.

Mylonas, D., Caparos, S., & Davidoff, J. (2022). Augmenting a colour lexicon. *Humanities and Social Sciences Communications*, *9*(1), 1–12. https://doi.org/10.1057/s41599-022-01045-3

Ocelak, R. (2016). "Categorical perception" and linguistic categorization of color. *Review of Philosophy and Psychology*, *7*(1), 55–70. https://doi.org/10.1007/s13164-015-0237-4

Peeples, D. R., & Teller, D. Y. (1970). Color vision and brightness discrimination in two-month-old human infants. *Science*, *189*, 1102–1103.

Peeples, D. R., & Teller, D. Y. (1978). White-adapted photopic spectral sensitivity in human infants. *Vision Research*, *18*, 49–53.

Pilling, M., Wiggett, A., Ozgen, E., & Davies, I. R. L. (2003). Is color "categorical perception" really perceptual? *Memory & Cognition*, *31*(4), 538–551. https://doi.org/10.3758/BF03196095

Plunkett, K., Hu, J.-F., & Cohen, L. B. (2008). Labels can override perceptual categories in early infancy. *Cognition*, *106*, 665–681. https://doi.org/10.1016/j.cognition.2007.04.003

Ramsay, J. O., & Silverman, B. W. (1997). *Functional data analysis*. Springer series in statistics. Springer New York. http://link.springer.com/10.1007/978-1-4757-7107-7

Regier, T., & Kay, P. (2009). Language, thought, and color: Whorf was half right. *Trends in Cognitive Sciences*, *13*(10), 439–446. https://doi.org/10.1016/j.tics.2009.07.001

Regier, T., & Xu, Y. (2017). The Sapir-Whorf hypothesis and inference under uncertainty. *Wiley Interdisciplinary Reviews: Cognitive Science*, *8*(6). https://doi.org/10.1002/wcs.1440

Roberson, D., & Davidoff, J. (2000). The categorical perception of colors and facial expressions: The effect of verbal interference. *Memory and Cognition*, *28*(6), 977–986.

Roberson, D., Davidoff, J., Davies, I. R. L., & Shapiro, L. R. (2004). The development of color categories in two languages: A longitudinal study. *Journal of Experimental Psychology. General*, *133*(4), 554–71. https://doi.org/10.1037/0096-3445.133.4.554

Roberson, D., Davies, I. R. L., & Davidoff, J. (2000). Color categories are not universal: Replications and new evidence from a stone-age culture. *Journal of Experimental Psychology: General*, *129*(3), 369–398. https://doi.org/10.1037/0096-3445.129.3.369

Roberson, D., Hanley, J. R., & Pak, H. (2009). Thresholds for color discrimination in English and Korean speakers. *Cognition*, *112*(3), 482–487. https://doi.org/10.1016/j.cognition.2009.06.008

Roberson, D., Pak, H., & Hanley, J. R. (2008). Categorical perception of colour in the left and right visual field is verbally mediated: Evidence from Korean. *Cognition*, *107*(2), 752–762. https://doi.org/10.1016/j.cognition.2007.09.001

Saji, N., Imai, M., & Asano, M. (2020). Acquisition of the meaning of the word orange requires understanding of the meanings of red, pink, and purple: Constructing a lexicon as a connected system. *Cognitive Science*, *44*(1), 1–23. https://doi.org/10.1111/cogs.12813

Sirois, S., & Jackson, I. R. (2012). Pupil dilation and object permanence in infants. *Infancy*, *17*(1), 61–78. https://doi.org/10.1111/j.1532-7078.2011.00096.x

Siuda-Krzywicka, K., Boros, M., Bartolomeo, P., & Witzel, C. (2019). The biological bases of colour categorisation: From goldfish to the human brain. *Cortex*, *118*, 82–106. https://doi.org/10.1016/j.cortex.2019.04.010

Skelton, A. E., Catchpole, G., Abbott, J. T., & Franklin, A. (2017). Biological origins of color categorization. *Proceedings of the National Academy of Sciences*, *114*(21), 5545–5550. https://doi.org/10.1073/pnas.1612881114

Teller, D. Y. (1979). The forced-choice preferential looking procedure: A psychophysical technique for use with human infants. *Infant Behavior and Development*, *2*(1), 135–153. https://doi.org/10.1016/S0163-6383(79)80016-8

Teller, D. Y. (1998). Spatial and temporal aspects of infant color vision. *Vision Research*, *38*(21), 3275–3282.

van Rij, J., Hendriks, P., van Rijn, H., Baayen, R. H., & Wood, S. N. (2019). Analyzing the time course of pupillometric data. *Trends in Hearing*, *23*, 1–22. https://doi.org/10.1177/2331216519832483

van Rij, J., Wieling, M., Baayen, R. H., & van Rijn, H. (2020). itsadug: Interpreting Time Series and Autocorrelated Data using GAMMs. R package version 2.4.

Wagner, K., Dobkins, K., & Barner, D. (2013). Slow mapping: Color word learning as a gradual inductive process. *Cognition*, *127*(3), 307–317. https://doi.org/10.1016/j.cognition.2013.01.010

Wagner, K., Jergens, J., & Barner, D. (2018). Partial color word comprehension precedes production. *Language Learning and Development*, 1–21. https://doi.org/10.1080/15475441.2018.1445531

Wass, S. V., Smith, T. J., & Johnson, M. H. (2013). Parsing eye-tracking data of variable quality to provide accurate fixation duration estimates in infants and adults. *Behavior Research Methods*, *45*(1), 229–250. https://doi.org/10.3758/s13428-012-0245-6

Witzel, C. (2016). New insights into the evolution of color terms or an effect of saturation? *i-Perception*, *7*(5). https://doi.org/10.1177/2041669516662040

Witzel, C., & Gegenfurtner, K. R. (2011). Is there a lateralized category effect for color? *Journal of Vision*, *11*(12), 1–25. https://doi.org/10.1167/11.12.16

Witzel, C., & Gegenfurtner, K. R. (2013). Categorical sensitivity to color differences. *Journal of Vision*, *13*(2013), 1–33. https://doi.org/10.1167/13.7.1.doi

Witzel, C., & Gegenfurtner, K. R. (2016). Categorical perception for red and brown. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(4), 540–570. https://doi.org/10.1037/xhp0000154

Yang, J., Kanazawa, S., Yamaguchi, M. K., & Kuriki, I. (2016). Cortical response to categorical color perception in infants investigated by near-infrared spectroscopy. *PNAS*, *113*(9), 2370–2375. https://doi.org/10.1073/pnas.1512044113

Younger, B. A. (1985). The segregation of items into categories by ten-month-old infants. *Child Development*, *56*(6), 1574–1583.

Yurovsky, D., Wagner, K., Barner, D., & Frank, M. C. (2015). Signatures of domain-general categorization mechanisms in color word learning. In *Proceedings of the 37th Annual Conference of the Cognitive Science Society*, (pp. 2775–2780).

Zhou, K., Mo, L., Kay, P., Kwok, V. P. Y., Ip, T. N. M., & Tan, L. H. (2010). Newly trained lexical categories produce lateralized categorical perception of color. *Proceedings of the National Academy of Sciences*, *107*(22), 9974–9978. https://doi.org/10.1073/pnas.1005669107

**How to cite this article:** Forbes, S. H., & Plunkett, K. (2023). Colour perception changes with basic colour word comprehension. *Developmental Science*, e13406. https://doi.org/10.1111/desc.13406