Supporting Information

Skelton et al. 10.1073/pnas.1612881114

SI Familiarization and Novelty Preference

Fig. S1 provides the average looking time at hue during the eight familiarization trials and the novelty preference scores for pairs of adjacently sampled hues (adjacent pairs) and larger hue pairs that straddle pairs where no novelty preference was found (large pairs). A repeated-measures ANOVA with hue (14 levels) and trial (8 levels) as factors revealed a linear trend for trial, $F(7, 1{,}148) = 53.965$, $P \leq 0.001$, $B = 3.29E+74$, on infant looking time during the familiarization phase. There was no interaction of trial and hue, $F(14, 164) = 1.28$, $P = 0.225$, $B = 0.002$.

SI Analysis of WCS Centroids

To investigate the hypothesis that infant color categories are organized around hues that are commonly central to lexical categories, we conducted an analysis using the WCS naming centroids [the number of times that the hue is at the center of a WCS lexical category when plotted in a perceptual color space (CIELAB)]. If infant color categories do capture the common centers of lexical categories revealed by the WCS, then the hue pairs that are categorically different for infants (indicated by pairs with novelty preference) should span hues which are infrequently at the centers of lexical categories (i.e., coincide infrequently with category centroids when counted across WCS languages). We investigated whether the combination of five pairs for which we find novelty preference optimally avoids category centroids [centroid frequencies were taken from Kay and Regier (8)]. We consider the 13 adjacent stimuli pairs from row G, hue columns (3, 6), (6, 9), (9, 12), (12, 15), (15, 18), (18, 21), (21, 24), (24, 27), (27, 30), (30, 33), (33, 36), (36, 39), (39, 2). The five pairs corresponding to infant novelty preferences are (3, 6), (12, 15), (21, 24), (30, 33), (36–39).

A simple means to quantitatively assess how well the infant preferences align with the WCS naming centroids is to assign a score to this particular configuration: the total number of centroids between the five stimuli pairs (excluding the stimuli column centroid counts). In this case, the score would be the sum of the centroid frequencies for hue columns (4, 5), (13, 14), (22, 23), (31, 32), (37, 38). We can then examine how well this particular configuration of 5 hue pairs (corresponding to the infant novelty preferences) ranks against any of the 1,287 possible combinations of 5 hue pairs, by computing the score of each of these configurations. To do so, we first computed the naming centroids over all 110 languages in the WCS, following Kay and Regier (8). We then explicitly computed the score for all 1,287 possible combinations of 5 hue pairs, using the centroid counts along the stimulus row that we sampled from (row G). We found that only 4.27% of the five pair combinations (55 of 1,287) resulted in a lower total centroid count than the five pairs for which infants had novelty preferences (i.e., more optimally captured hues that are not central to lexical categories).

Additionally, we used the peak centroid counts for each hue at any lightness level so that the centroid peaks seen in Fig. S2 are captured in the analysis. We found that only 3.26% (42 of 1,287) of the combinations resulted in a lower total centroid count than the five pairs for which infants had novelty preferences. For this peak centroid analysis, the best five-pair combination that had the lowest centroid count included four of the five pairs for which infants showed a novelty preference, and only the blue–green pair differed (the best combination included the pair adjacent to and bluer than the blue–green pair for which infants showed a novelty preference).

SI Discrimination and Novelty Preference

The definition of categorization is that "discriminable stimuli are treated equivalently" (29). One possibility is that a lack of a novelty preference could indicate that infants cannot actually see the difference between colors rather than indicate that infants can discriminate the colors, but treat them equivalently in memory. We predicted, on the basis of a study of infant chromatic thresholds (34), that all chromatic differences should be discriminable for infants because all chromatic differences were larger than the average chromatic discrimination threshold at 4-mo (threshold estimated to be 21ΔE units in CIELAB color space) (34): the smallest difference in the current study was 28ΔE and the average was 62.38ΔE for the smaller pairs and 109.47 for the larger pairs. However, there are difficulties in inferring discriminability from other studies that used different stimuli (e.g., Knoblauch's estimates were for chromatic differences from neutral along protan, deutan, and tritan axes, and the threshold estimates also had a good deal of variability). In addition, even if all chromatic differences were above chromatic thresholds, it is possible that those stimuli with novelty preferences are more perceptually dissimilar. Therefore, to further understand the relationship between discrimination, perceptual similarity, and novelty preference, we took two approaches. First, we investigated the relationship between infants' novelty preferences and the similarity of colors by testing whether the size of color differences in a perceptual color space and adult similarity ratings of the colors predicted infants' novelty preferences. Second, we made additional measurements that directly tested infant discrimination of four of the color pairs (two with novelty preferences and two without) to check that colors that failed to elicit a novelty preference could be discriminated in other contexts, and to assess the relationship between discriminability and novelty preference.

Stimulus Positions in a Perceptually Uniform Color Space. A series of regressions found that differences in CIE hue, CIE chroma, and Euclidean difference in CIELAB perceptual color space (Fig. S3) as predictors of novelty preference did not predict novelty preference (largest $R^2 = 0.072$, smallest $P = 0.31$, $B = 3.28E-10$). In addition, a sample of 40 adults (5 males, mean age = 21.17 y; SD = 1.37) rated the similarity of hue pairs twice using a line rating scale, and adult similarity ratings did not predict infants' novelty preference ($R^2 = 0.083$, $P = 0.28$, $B = 1.13–09$).

This analysis indicates that the color pairs that infants do not distinguish in recognition memory are not the smaller chromatic differences in adult perceptual color space that adults find more similar than other pairs. However, one possibility is that infant perceptual color space at 4–6 mo is different to that of adults (even in adults CIELAB is not perfectly uniform for large chromatic differences) (53). A difference in infant and adult perceptual color space would need to be radical to be able to account for the pattern of novelty preferences in our data because there was even no novelty preference for pairs with very large CIELAB differences. However, a radical difference between infants and adults is perhaps unexpected because infants are known to be trichromatic by at least 3-mo (e.g., ref. 35), mean adult isoluminance is a good approximation of infant isoluminance (54), and based on chromatic thresholds on protan, deutan, and tritan axes, their perceptual color space could be predicted to be similar to that of adults, albeit with poorer sensitivity (34).

Measurements of Discrimination and Perceptual Similarity for Stimulus Pairs. We tested discrimination and perceptual similarity using a target-detection task developed by Franklin, Pilling, and Davies

(25). In our target-detection task, a colored target is seen on a colored background and eye movements are recorded with an eye-tracker to measure how well infants can discriminate the target from the background. Low-pass luminance noise was added so that the target could only be detected on the basis of the chromatic difference of target and background and not on the basis of luminance [colors were isoluminant for the average adult observer, which is a good estimate of infant isoluminance (54)]. The task was also made to be gaze-contingent, such that a visual and auditory reward was played when the target was fixated and the next trial commenced automatically following target fixation (as in ref. 55). Four stimulus pairs were tested from green–yellow and purple–red regions. For both regions, two adjacent pairs were sampled, one pair with a novelty preference and one without. We checked whether these pairs, which include two of the smallest chromatic differences (in CIELAB), could be discriminated in the context of the target detection task, and compared the perceptual similarity of the color pairs.

Of course, if infants can discriminate the colors in the context of the target detection task it does not necessarily indicate that they can be discriminated when seen in the context of the novelty preference task, as the spatial characteristics of the task differ. Because the stimulus pairs in the target detection are abutting as figure and ground, this could lower discrimination thresholds on the task relative to the test phase of the novelty preference task where the two different stimuli were surrounded by a neutral ground. Unfortunately, in infants it is not possible to measure discrimination of two stimuli using looking measures when the stimuli are shown side by side on a neutral ground, as in the novelty preference task, but without a familiarization phase. Infants may look longer at one stimulus than the other, which would indicate discrimination, but may not have a looking preference and still be able to discriminate them (e.g., be able to discriminate blue and red but look at them equally). However, the target-detection task does provide a measure of relative perceptual similarity across the color pairs. If novelty preferences are based on how different colors look rather than their categorical relationship, then the pairs for which there were novelty preferences should have greater target fixation than the pairs with no novelty preference.

SI Methods

Participants. Twelve 4- to 6-mo-old infants took part, with two infants excluded because of infant fussiness. The final sample (eight males) had a mean age of 22.63 wk (SD 2.65). All infants had a birth weight greater than 2,500 g and no known visual or neurological conditions.

Apparatus and Stimuli. There were four stimulus pairs, sampled from the main experiment from green–yellow (2.5Y–10Y; 10Y–7.5GY) and blue–purple regions (2.5P–10P; 10P–7.5RP). One other stimulus (7.5B) was chosen to give a measure of chance performance (further explained below). The luminance and chromaticity coordinates (CIE x, y, Y, 1931) of the Munsell stimuli and the gray background for the novelty preference task were measured with a photospectrometer (spectrascan PR6500) in the viewing booth and under the same lighting conditions as the novelty preference task. These x, y, Y were then rendered on a calibrated 22-inch Mitsubishi DiamondPlus 2070SB Diamondtron CRT monitor with a resolution of $1,600 \times 3 \times 1,200$ pixels, 24-bit color resolution, and a refresh rate of 100 Hz. Stimuli were displayed via a PC-driven Cambridge Research Systems ViSaGe MKII Stimulus Generator. Stimuli were shown as a colored circular target (which subtended a visual angle of 6.9°) on a colored background that filled the entire screen, with low-pass luminance noise (12 cd/m^2). The experiment took place in a blacked-out booth with the monitor being the only source of light. Eye movements were recorded with an Eyelink 1000 eye-tracker (SR-Research), placed immediately in front and below the monitor, and the central point of the monitor screen was at the participants' eye-level at a distance of 35 cm. Participants were sat in a car seat that was fixed to a stable chair.

Design and Procedure. Infants were shown a cartoon while the eye-tracker was focused on their eye, and this was followed by a four-point calibration. On each trial, the colored target was displayed simultaneously with the colored background up until target fixation but not longer than 2,250 ms. Fixation was defined as 160 ms of continuous looking at one point, as in (ref. 55). If the target was fixated, the program automatically displayed a smiley schematic face in the same location as the target (of the same size, with eyes and mouth defined in gray) and a simple melody was played for 500 ms. Following this, or after 2,250 ms if there was no target fixation, the next target and background were presented, with the target location constrained with the target positioned at a random location 6.9° of visual angle from the infants' initial point of fixation (55). If infant gaze strayed from the screen, the trial was discounted, a looming and contracting black and white attention getter was displayed centrally, and the next trial was begun once the attention getter was fixated. In addition to the four stimulus pairs from the novelty preference task, there was a condition where the target and background were the same color (7.5B) to estimate chance performance on the task. The allocation of each color in a stimulus pair as target or background was counterbalanced, and trial order was randomized. Trials continued until infant looking at the screen waned or until a maximum of 120 trials had been completed.

SI Results and Discussion

The percentage of trials where the target was fixated was calculated for the four green–yellow and purple–red stimulus pairs (hit rate), and the condition where target and background were identical (guess rate). Fig. S4 gives the guess rate and the hit rates for the four pairs (left side of Fig. S4) with the novelty preferences for the same pairs from the main experiment for comparison (right side of Fig. S4).

Bayesian paired-samples t tests (with a Cauchy prior of 0.707) revealed that for all four pairs the target was fixated at a rate greater than chance, meaning that all four color pairs were discriminable: 7.5GY–10Y, $t(9) = 3.13$, $P = 0.01$, $B = 5.32$; 10Y–2.5Y, $t(9) = 2.77$, $P = 0.02$, $B = 3.31$; 2.5P–10P, $t(9) = 8.04$, $P < 0.001$, $B = 1081.83$; 10P–7.5RP, $t(9) = 6.11$, $P < 0.001$, $B = 175.26$. Moreover, fixation was not more likely for the two pairs that elicited novelty preferences (7.5GY–10Y and 10P–7.5RP, mean = 42.18, SD = 18.41) than the two pairs that did not (10Y–2.5Y and 2.5P–10P, mean = 41.98, SD = 5.85), $t(9) = 0.09$, $P = 0.93$, $B = 0.31$. These findings clearly show that color differences that failed to elicit a novelty preference are discriminable by infants in other contexts. Even if the context of the target-detection task enhances discriminability, novelty preferences are unlikely to be simply related to perceptual similarity because targets for color pairs with greater novelty preference were not always fixated at a greater rate than color pairs with no novelty preference.

In sum, we show here that infants' novelty preferences are unlikely to be a result of discriminability or perceptual similarity as: (*i*) all color differences were above estimated average chromatic thresholds; (*ii*) novelty preferences do not relate to chromatic differences in perceptual color space; and (*iii*) chromatic differences that failed to elicit a novelty preference are discriminable in other contexts and not always detected more readily than chromatic differences with a novelty preference. We therefore suggest that infants' pattern of novelty preference indicates that infant color memory is governed by the categorical relationship of the colors.

The present study demonstrates that color categories affect infant hue memory. Another question is whether color categories themselves affect perception (categorical perception) (25). That question is different from the one addressed in the target detection experiment outlined above. The experiment above seeks to obtain a measure of perceptual similarity to see whether perceptual similarity rather than categories can account for infants' novelty preferences. An experiment that tested categorical perception would need to instead equate same- and different-category chromatic differences in discrimination and see whether categories affect perceptual similarity. Franklin et al. found that colored target detection was faster when on different- than same-category colored backgrounds, when the chromatic differences were equated in CIE perceptual color space (25). However, as discussed above, it is an assumption that adult and infant perceptual color spaces are similar, and equating stimuli in discrimination rather than perceptual similarity would be more logical for testing categorical perception. One approach that may clarify whether color categories affect infant perceptual similarity would be to equate same- and different-category color differences in the number of just-noticeable differences, and then test whether the categorical status affects target detection when chromatic differences are suprathreshold. This approach was adopted by He et al. (47) to investigate the time course of color category effects in adults, and when colors were equated in just-noticeable difference, color categories affected postperceptual processing around 200 ms from stimulus onset. Now that the current investigation has identified infants' categorical distinctions, the approach taken by He et al. can be applied to infants.

SI A Priori Preference

Infants have a priori preferences for looking longer at some hues (e.g., reds and blues) than others (e.g., yellow–greens) (e.g., refs. 56–59). The present study controlled for such a priori preferences by counterbalancing across infants which stimulus was novel for each pair of stimuli. Because of this counterbalancing, preference for a certain hue (e.g., blue) is unlikely to account for novelty preference for a stimulus pair (e.g., blue and green) because half of the infants saw a given hue as the novel color and half as the familiar hue; when novelty preference is averaged across infants the effect of any a priori preferences should cancel out (60). Mean novelty preferences were in fact highly similar irrespective of which stimulus in each pair was novel and which was familiar: the discrepancy in novelty preference was on average only 5%. In addition, a priori preferences are unlikely to account for novelty preferences because a priori hue preference in infants varies smoothly with hue, and therefore two similar hues are unlikely to elicit a large difference in looking times (59, 60). An analysis was conducted to confirm that a priori preferences could not account for variation in novelty preference across stimulus pairs. The average looking time at each hue during the familiarization phase was used as an index of a priori preference. Looking time during the familiarization phase is a valid index of a priori preference, as prior research has established that infants' a priori preferences for hues are comparable whether hues are presented singly (as in the familiarization phase) or as pairs (as in the test phase) (61). The average familiarization looking time at a stimulus (during either the first trial or across all familiarization trials) did not predict infants' novelty preference for that stimulus, largest $R^2 = 0.004$, smallest $P = 0.38$, largest $B = 0.14$. Similarly, the difference in average familiarization looking time for the two stimuli in each pair did not predict the variation in novelty preference across pairs, $R^2 = 0.009$, smallest $P = 0.19$, largest $B = 0.22$.

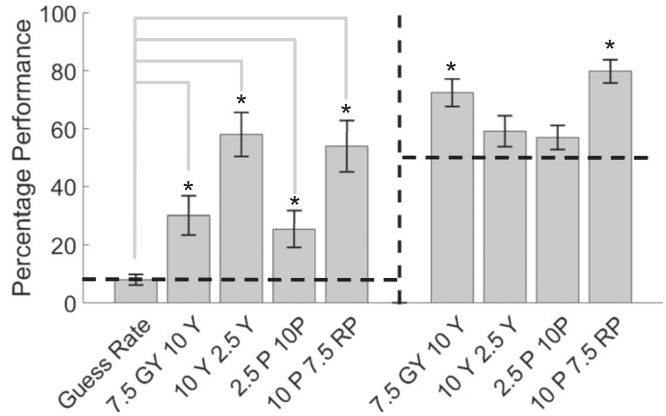

Fig. S1. The average percentage of targets fixated (left side of figure) for the four pairs tested in the target-detection task, with novelty preferences from the main experiment for comparison (right side of figure). The dashed line in each panel indicates chance performance (also indicated by the guess rate in the target fixation panel). Asterisks indicate pairs for which the evidence was in favor of H1 compared with chance.

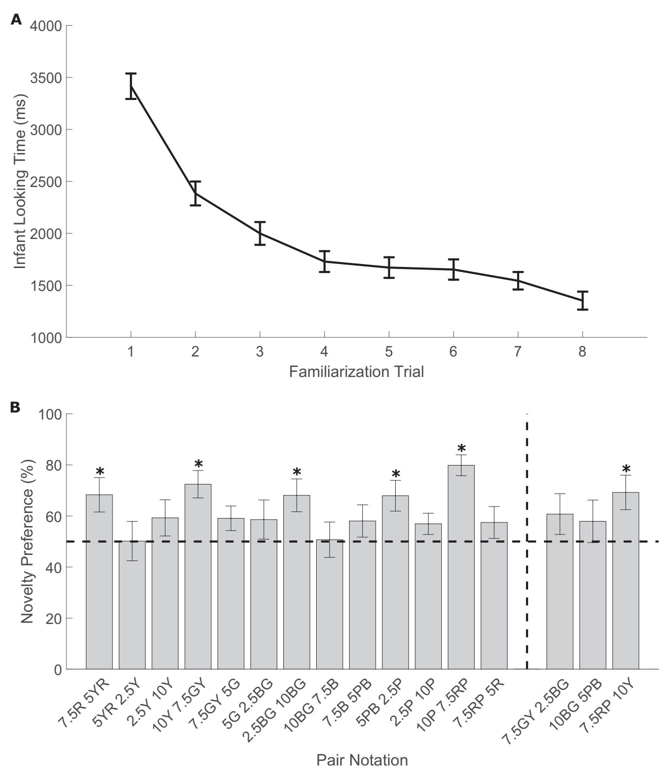

Fig. S2. (*A*) Infant looking time during the familiarization trials averaged across hues (±1 SE). (*B*) Novelty preference percentages (±1 SE) for pairs of adjacently sampled hues (pairs on left of vertical dashed line in figure) and larger hue pairs which straddle smaller pairs (pairs on right of vertical dashed line in figure). Asterisks indicate novelty preferences that produce a Bayes factor in support of H1 relative to H0 ($B > 3$), with equal looking at novel and familiar hues indicated by the dashed line at 50%. Munsell codes of hues of each pair are given.

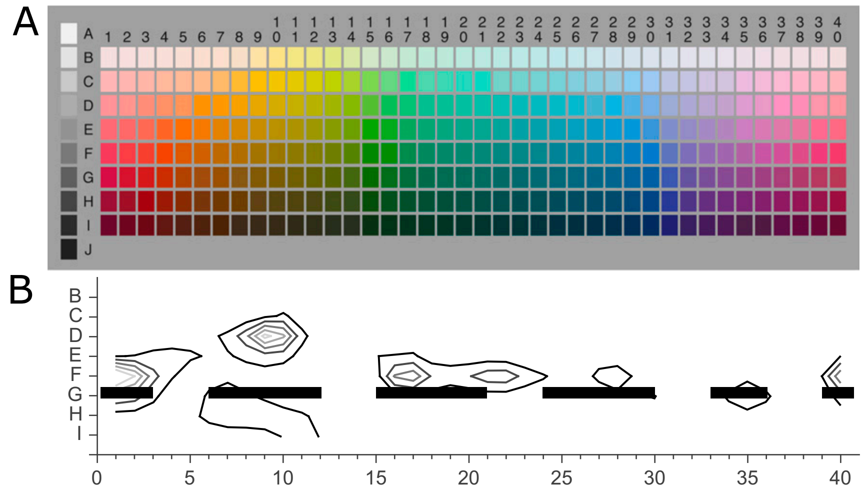

Fig. S3. The full WCS stimulus grid and centroid analysis. (*A*) WCS stimulus grid for 40 hues at 8 lightness levels. Stimuli in the current experiment were sampled from row G. (*B*) Centroid analysis from Kay and Regier (8). The black and gray contour lines indicate the number of speaker centroids falling at that point in the stimulus grid from WCS data, with the outermost contour representing 100 centroids and each subsequent inner contour representing an increment in 100 centroids. Infant novelty preferences are indicated by the gaps in the thick black horizontal line at row G.

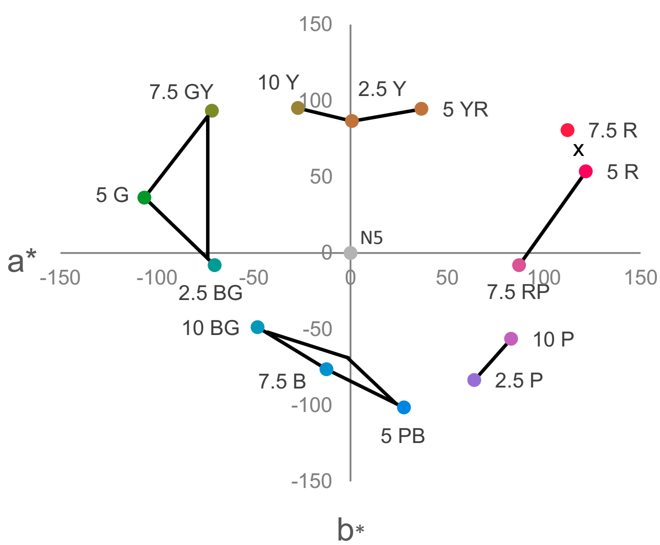

Fig. S4. Stimuli plotted in CIELAB perceptual color space (a*,b*). Stimulus pairs for which there was no novelty preference are indicated with black lines joining the stimuli, and pairs where there was a novelty preference are indicated by the absence of these lines. The cross between 7.5R and 5R indicates a pair that was not tested. Euclidean distances in this space do not predict infants' novelty preference.

Other Supporting Information Files

Table S1 (DOCX)