Supplementary Information

Farnsworth-Munsell 100-Hue Test Results

Figure S1 shows the FM100 total error scores of two participants, one who performed below
average and one who performed above average.

Chromatic Contrast Discrimination Threshold Test (CCDT)

Table S1. Participant group demographics, showing only participants whose results were included in
the CCDT analysis. Chronological age is shown in years. VIQ and NVIQ are shown as standardised
scores. In all groups IQ was assessed using the WISC Fourth Edition. Standard deviations are shown
in brackets.

<table>
<thead>
<tr>
<th>Group</th>
<th>Chronological Age</th>
<th>VIQ</th>
<th>NVIQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>TD6-7y (n=12)</td>
<td>6.38 (0.17)</td>
<td>113.33 (12.17)</td>
<td>107 (12.45)</td>
</tr>
<tr>
<td>TD8-9y (n=18)</td>
<td>8.96 (0.31)</td>
<td>121.94 (11.87)</td>
<td>112.83 (13.61)</td>
</tr>
<tr>
<td>ATY Child (n=16)</td>
<td>12.15 (2.42)</td>
<td>71.25 (22.05)</td>
<td>85.13 (19.59)</td>
</tr>
</tbody>
</table>

Stimuli: The contrast of the arrow stimulus with respect to the background was calculated in cone-
opponent contrast coordinates following the formulae given in Eskew et al. (1999). The origin of this
coordinate space is given by the cone excitations to a reference white surface (L₀, M₀ and S₀), in this
case, the uniform grey background, using the Smith-Pokorny (1975) cone fundamentals.¹ Cone
excitation values for the arrow are defined with respect to this origin by the ratios: \( \Delta L = (L - L_0)/L_0, \)
\( \Delta M = (M - M_0)/M_0 \) and \( \Delta S = (S - S_0)/S_0, \) where L, M, and S are the cone excitations to the arrow
stimulus. The cone-opponent contrast coordinates of the arrow stimulus are then calculated for
each of three axes:

\[ LUM = 0.78 \Delta L + 0.37 \Delta M; \]
RG = [0.7*ΔL – 0.72*ΔM + 0.02* ΔS];

BY = – [0.55*ΔL + 0.25 ΔM – 0.8ΔS];

where LUM is the luminance axis, RG is the “red-green” axis and BY is the “blue-yellow” axis.

For each half of each axis, a set of displayable colours (numbering on average 75) were calculated whose colour differed from the uniform background only along that axis, in the smallest increments achievable for the given display device. For example, for the “yellower” direction of the BY axis and the 10bit display, all ~65 stimuli had effectively the same LUM and RG coordinates as the grey background while their BY coordinates varied systematically from the background BY coordinate in steps equal to (on average) 0.3142 ΔE_{u\*v\*}, up to a maximum difference of approximately 20 ΔE_{u\*v\*}. (The unit ΔE_{u\*v\*} is defined as a just-noticeable-difference in the perceptually uniform CIE Lu*v* space.)

**Staircase Procedure:** Six individual staircases, one for each direction of each colour axis (i.e. each half-axis), were completed by each observer. Each staircase drew its stimuli from one of the colour sets calculated as described above. Each colour axis was tested with a separate block of trials, in which the two half-axis staircases were interleaved; e.g. for the BY colour axis, the “bluer” and “yellower” staircases were interleaved. Within each half-axis staircase, the arrow colour was varied in 6 decreasing step sizes ranging from approximately 10 ΔE_{u\*v\*} to 0.3 ΔE_{u\*v\*} (for the 10-bit experiment; step-sizes ranged from approximately 10 ΔE_{u\*v\*} to 0.8 ΔE_{u\*v\*} for the 8-bit experiment) in a one-up/two-down procedure. The steps were selected in terms of the indices of the colour stimuli in the colour sets described above, with the smallest step in each case corresponding to an index step-size of 1.
Figure S2 shows an illustration of the stimuli used in the CCDT and the temporal sequence of events over one trial.

**Comparison of 8-bit and 10-bit display systems for CCDT**

In this paper, the CCDT was used with two different graphics boards/display systems (8-bit and 10-bit), due to changes in equipment availability over the course of the study. Although the 10-bit system is preferable for widespread use because of its higher chromatic resolution, the two systems are equally effective for the populations studied here. A control experiment in 5 adult individuals, who performed the test on both 8-bit and 10-bit systems in the same day, revealed a significant positive correlation between thresholds on the two systems ($r = 0.88; p < 0.001$). All thresholds obtained on the two systems are also comparable to thresholds reported for other standardised chromatic discrimination tests, e.g. the Colour Assessment and Diagnosis (CAD) test and the Cambridge Colour Test (CCT), in both absolute magnitude and variation across chromatic direction.

Note that because thresholds vary across colour space, with both the background chromaticity and luminance, it is not appropriate to make exact comparisons.

**Regression Model**

Further regression models were run on data from all participants (child and adult groups), using TES as a dependent variable and NVIQ, VIQ, Chronological Age (months), and Development as predictors, excluding CCDT (Backwards entry method). Three iterations of model yielded the following predictions:

1) \( R^2 = 64.9 \% \), all predictors included
2) \( R^2 = 64.9 \% \), Development only removed
3) \( R^2 = 64.3 \% \), Development and VIQ removed (only NVIQ and CA remaining)
NVIQ is always the highest predictor, but with slightly lower weighting than in results from Model 2 presented in Table 2 of the main paper. A similar final regression model run for the child groups only (with NVIQ and Chronological Age only as predictors after Development and VIQ have been removed) yields $R^2$ of 47.1%. For the adult groups only, the final regression model explains 66.9% of the variance (with NVIQ only as predictor, after Development, VIQ and Chronological Age have been removed).

All models explain more variance when not including CCDT measures. For data from the child groups only, CCDT thresholds (converted to logarithms) are a significant predictor when they are the only predictor but explain much less of the variance (23.4%) than the NVIQ-only model (39.5%). For data from both the child and adult groups, the NVIQ only model predicts slightly more variance (41.4%). The adult only NVIQ only model predicts much of the variance (66.9%).

Regression models were also run using the CCDT threshold from each colour axis separately with its corresponding FM100 partial error score (PES), for example, using only the Red-Green PES and Red-Green CCDT thresholds within one regression analysis and the Blue-Yellow PES and Blue-Yellow CCDT thresholds in a separate analysis. The final Red-Green model explained 44% of the variance (final model predictors: Development, Red-Green CCDT threshold and NVIQ). The Blue-Yellow model explained more of the variance (61.3%, final model predictors: Chronological Age, Blue-Yellow CCDT threshold and NVIQ) than the Red-Green model. Despite the difference in overall variance explained between the Red-Green and the Blue-Yellow model, NVIQ was still the highest loading predictor in both models.

A final regression analysis was done using Development, Chronological age, NVIQ, Red-Green CCDT threshold and Blue-Yellow CCDT threshold as predictors and FM100 TES as a dependent variable. Five iterations of the model yielded the following predictions:

1) $R^2 = 54.8\%$, All variables included
2) $R^2 = 54.8\%$, Development removed
3) $R^2 = 52\%$, Blue-Yellow CCDT threshold removed
4) \( R^2 = 47.8\% \) Red-Green CCoD threshold removed

5) \( R^2 = 43.3\% \) Chronological Age removed (Only NVIQ remaining).