Vision Research xxx (xxxx) 1–13

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

Colour category constancy and the development of colour naming

Christoph Witzel [a], Zoe Flack [b], Emma Sanchez-Walker [c], Anna Franklin [c]

[a] School of Psychology, University of Southampton, United Kingdom
[b] School of Humanities and Social Science, University of Brighton, United Kingdom
[c] The Sussex Colour Group, School of Psychology, University of Sussex, Falmer, Brighton BN1 9RH, United Kingdom

ARTICLE INFO

Keywords:
Category development
Colour categorisation
Colour constancy
Colour perception
Language acquisition

ABSTRACT

In this study, we investigated the processes of coordination, adaptation, and calibration during the development of colour naming and colour constancy, and we tested whether colour term knowledge is related to colour constancy. We measured category membership and prototypes with 163 Munsell chips in preschool children (3- to 4-year-old) under neutral, green, and red illuminations, and compared their results to those of adults. We introduced an index of colour term maturity based on the similarity of children's colour term use to adults, and a colour category constancy index that quantifies the variation in colour categorisation that is specific to illumination changes. Results showed that illumination changes affected children's consistency of colour categorisation, but only to a small extent. However, colour term maturity and illumination-specific effects on consistency strongly varied in this age range. Correlations between colour term maturity and illumination-specific consistency indicated that colour constancy increases with colour term acquisition; but those results depended on the type of illumination changes (between neutral, green, and red). Together, our findings suggest that children progressively fine-tune and recalibrate the meaning of colour terms through processes of coordination and adaptation that are also involved in the calibration of colour constancy.

1. Introduction

In this study, we investigated the processes of coordination, adaptation, and calibration during the development of colour naming and colour constancy. In colour naming, speakers need to coordinate with each other how to use colour terms to successfully communicate colours (Steels & Belpaeme, 2005). Colour constancy is the ability to recognise colours across illuminations (Smithson, 2005; Hurlbert, 2007; Foster, 2011). To achieve colour constancy, observers need to adapt to the variation of colours across illuminations in their visual environment (Witzel & Gegenfurtner, 2018b).

The challenges of naming and constancy are common to most perceptual domains, such as object recognition and constancy, size constancy etc (e.g. Gelman & Meyer, 2011; Owen & Barnes, 2021; Yang, Kanazawa, Yamaguchi, & Motoyoshi, 2015). Yet, the case of colour naming provides a particular opportunity to study language acquisition because of the discrepancy between colour perception and lexical colour categories. On the one hand, human observers are able to perceive colours continuously along 3 dimensions; on the other, colour terms categorise the multitude of perceivable colours in a few colour categories that collapse the three perceptual dimensions (Fig. 4 in Witzel & Gegenfurtner, 2018b). The lack of a correspondence between perceptual dimensions and colour terms provides the opportunity to identify the effects of language acquisition in behavioural measurements of categorisation.

The acquisition of colour categories involves two sources of variation: the variation of candidate colours that could be denoted by a colour term as well as the variation of colours across lighting conditions. When children acquire language, they not only need to learn the words that designate a single object or attribute; they also need to learn the linguistic distinctions that allow drawing a boundary between what belongs and what does not belong to the category of a particular word (e.g. Gelman & Meyer, 2011). For example, they need to learn that the colour term "green" may refer to a large range of different shades of green (Fig. 4 in Witzel & Gegenfurtner, 2018b). At the same time, children need to learn that the same objects and attributes may change depending on external conditions, hence making it necessary to develop perceptual constancy. For example, lighting changes in everyday life, e.g., from direct sunlight to shadows or to indoor lighting. The sensory colour signal of the same greenish surface may strongly differ under direct sunlight, shadow, and under a tungsten light bulb (the sensory colour signal is the colour information at the first stage of colour processing as reflected, for example, by cone excitations or Tristimulus Values).

Children need to develop colour constancy to be able to communicate about colours through colour terms in a changing environment. There are a few studies that suggest that children have a certain degree

https://doi.org/10.1016/j.visres.2021.05.008
Received 20 August 2020; Received in revised form 19 May 2021; Accepted 19 May 2021
0042-6989/© 2021

of colour constancy before the age at which they acquire colour categories. Previous studies suggested that infants have colour constancy, at least partially, at the age of 4 months, but not before (Dannemiller, 1989; Dannemiller & Hanko, 1987; Yang, Kanazawa, Yamaguchi, & Kuriki, 2013). This is in line with the finding that children at this age are already sensitive to local contrast, which is a mechanism that supports colour constancy (Pereverzeva & Teller, 2009). It is also consistent with the findings that 4-month-old infants begin to have lightness constancy (Chien, Bronson-Castain, Palmer, & Teller, 2006; Granrud, Corrow, Gilchrest, Mathison, & Yonas, 2011).

While these studies suggest that some mechanisms of colour constancy might already be in place early in childhood, they do not imply that colour constancy is fully developed when children begin to acquire colour terms. Colour constancy may be explained to some extent by fixed sensory mechanisms of chromatic adaptation, which establish a *constant appearance* of surface colours across illumination changes. However, colour constancy also requires *inferential colour constancy*, i.e., unconscious inferences about illumination conditions that go beyond the effects of chromatic adaptation (Smithson, 2005; Foster, 2011; Witzel & Gegenfurtner, 2018b). In particular, the sensory colour signal does not contain all information necessary to achieve colour constancy. Due to metamer mismatching, the same colour signal may change in very different ways across illuminations depending on the wavelength spectra of the surfaces and illuminations at the source of the colour signal (Logvinenko, Funt, Mirzaei, Tokunaga, & Osorio, 2015; Cohen & Kappauf, 1982). This implies that human observers might need to learn how surfaces vary across illuminations in their visual environment (Witzel, van Alphen, Godau, & O'Regan, 2016; Witzel & Gegenfurtner, 2018b).

Colour constancy and colour categorisation might be subject to similar processes of learning and adaptation. In both cases, the child needs to understand that colours that look quite different need to be considered as equal, either in terms of the same colour term or in terms of the same surface. In the case of naming, a colour term refers to different shades of colour, e.g., "green" denotes various shades of green. In colour constancy, the same surface may correspond to different shades of colours depending on the illumination. We will refer to the idea that the same colour information refers to varying colour signals as *multivalence*. Due to multivalence, the child needs to learn which range of colours to expect when a certain colour term is communicated, or when a certain surface needs to be identified under other illuminations. Communicating colours across illuminations in everyday life situations may involve both sources of variation simultaneously, the variation in the use of a colour term for different colours and the variation in illumination. Colour constancy and colour categorisation might therefore be related during colour term acquisition because they both rely on the children's understanding of colour multivalence. A recent study provides evidence for this idea, showing that children with better colour term knowledge are also better at identifying colours across illuminations (Rogers, Witzel, Rhodes, & Franklin, 2020).

In addition, previous research on computational modelling and colour naming in adults suggests a link between colour constancy and categorisation. It has been argued that the colour signals corresponding to category prototypes are more predictable across illuminations (Philipona & O'Regan, 2006; Vazquez-Corral, O'Regan, Vanrell, & Finlayson, 2012). Further evidence for a relationship between colour constancy and colour naming comes from studies on colour category constancy in adults (Olkkonen, Hansen, & Gegenfurtner, 2009; Olkkonen, Witzel, Hansen, & Gegenfurtner, 2010). *Colour category constancy* ("categorical colour constancy" in those previous studies) is the stability of colour categories across illumination changes. Those studies found that the *colour category consensus* is strongly correlated with colour category constancy, where the consensus is the stability of colour categories across observers. In particular, category membership of colours in the category centres were perfectly stable, and only colours close to the category boundaries varied considerably. This was similar for the variation across observers, as for the variation across illuminations, hence the strong correlation between both. Prototypes, being close to the centre of categories, were almost perfectly stable (see also, Douven, Wenmackers, Jraissati, & Decock, 2017). Consequently, prototypes could be *anchors* for colour naming, i.e., points of stability in perceptual colour space that provide a reference for naming other colours whose category membership changes more strongly across illuminations and observers.

However, illumination induced shifts of colour signals are more likely to affect category membership of colours close to the category boundary. As a result, illumination induced shifts in category membership are likely to coincide with the low consensus of category membership at the boundary. This could produce the correlation between colour category constancy and consensus (Olkkonen et al., 2009, 2010) without any direct relationship between colour constancy and categorisation. In addition, other studies have not found a relationship between colour categories and perceptual colour constancy in adults and contradicted the idea that prototypes could act as perceptual anchors (Weiss, Witzel, & Gegenfurtner, 2017; Witzel et al., 2016).

It is completely open how children map colour categories across different illuminations during development. Is children's categorisation affected by illumination changes? This would indicate that failures of inferential constancy add to the difficulties of colour term acquisition by increasing the multivalence of colour terms. Or are children's categories similarly immune to illumination changes as adult categories? This would show that the main difficulty in colour term acquisition consists of figuring out the location of colour categories in perceptual colour space independent of illumination changes.

In this study, we measured the constancy of colour categories at the age of colour term acquisition. We employed a large sample of colours that allowed us to map colour categories across all three perceptual dimensions, hue, lightness, and saturation. We examined whether the development of colour categories is contingent upon the constancy of categories across illuminations. To this end, we separated computationally the variation of colour categories that is specific to illumination changes from inconsistencies in categorisation that are unrelated to illumination changes. In this way, we established the degree of colour constancy in children, and compared it to the level of colour constancy in adults.

If colour constancy is not yet fully developed at the age of colour term acquisition, children should have a considerably lower degree of colour constancy than adults. Most importantly, we tested whether there is a relationship between the level of colour constancy and the maturity of colour term use in those children. Here, we call *colour term maturity* a child's ability to use basic colour terms like adults, and we assess colour term maturity as the similarity in colour term use between children and adults (Witzel, Flack, & Franklin, 2013). This measure of colour term maturity focused on basic colour terms because these terms are used with high consensus among adults (e.g. Berlin & Kay, 1969; for review Witzel, 2018a). In free, unconstrained naming, there is no clear reference of maturity because even adults greatly vary in their use colour terms (e.g. Lindsey & Brown, 2014). We distinguish colour term maturity from *naming consistency*, which refers to how consistently a child uses the same colour terms for the same colours. Naming consistency reflects each child's ability to categorise colours, no matter whether the categories or the colour terms used to denote them correspond to those of adults or not. Preliminary results have been presented at a conference (Witzel, Sanchez-Walker, & Franklin, 2013).

2. Method

Fig. 1 illustrates the set-up of the main experiment. In sum, we asked preschool children to name a sample of 163 Munsell chips. The task was made into a fun game: Children were asked to find coloured

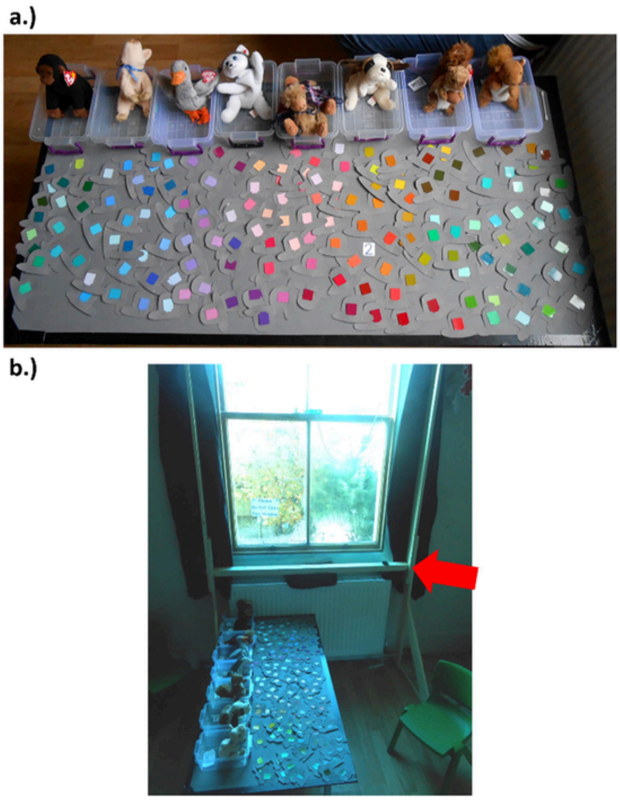

Fig. 1. Experimental set-up. a.) Colour chips in hat-shaped chip holders and toy animals used in the naming game. b.) Set-up under green illumination. The red arrow indicates the wooden frame with the filter.

items of clothing for toy animals. There were 8 toy animals corresponding to the 8 chromatic basic colour terms. The chips were put into grey chip holders in the shape of clothing items, such as the little hats in the photo of Fig. 1a. For each animal, children were asked to find all the items that had the colour the animal wants, for example all the red hats for the duck. This was done twice under neutral daylight, and once under red, and green illumination.

2.1. Child participants

To assess category development, we investigated an age range of 36–50 months, at which children just know all or almost all basic colour terms (cf. Fig. 6 in Pitchford & Mullen, 2002), but do not yet categorise all colours like adults. The sample consisted of 27 children, 10 girls and 17 boys. Their age varied between 36.6 months and 52.9 months, with a median age of 41.5 months (an age distribution is provided in Fig. S1 of the supplementary material). Due to dropouts, not all children completed all conditions. Five of them did only one neutral condition. This precluded them from the below analyses, which required comparisons across measurements. In addition, some children did not complete the prototype choices due to a lack of motivation. An overview of all participants and the conditions they completed is provided in Table 1.

Before the main experiment, we assessed red-green colour deficiencies with the non-numerical (winding lines) isochromatic Ishihara plates 38 (for practice), 37, 36, and 26. In case of doubt, we double-checked with corresponding plates 35, 34, and 27 (Ishihara, 2004). None of the children had colour deficiencies according to this test.

We also assessed children's colour term knowledge with a naming and comprehension tests adapted from Pitchford and Mullen (2002). The plates we used in these pre-tests can be found in Figs. S4-S5 of the supplementary material. All children were able to name 4–8 and to comprehend between 3 and 8 of the eight chromatic basic colour terms in the naming pre-test. This was necessary to make sure that children could accomplish the categorisation task at all. Ethical approval for this study was obtained from the Sciences and Technology Cross-Schools Ethics Committee at University of Sussex.

Table 1
Overview of participants. Rows correspond to participant groups. Columns refer to illumination conditions. Neutral2 is the repeated measurement under daylight without filter. Adults provided data for 450 chips including prototypes; children only for the subset of 163 maximally saturated Munsell chips. "x2" refers to the 5 German adults having done those conditions twice, once with full and once with reduced local contrast. In parenthesis are numbers including dropouts.

	neutral	green	red	neutral2
ADULTS				
British	10	–	–	–
German	5x2	5x2	5x2	5x2
CHILDREN				
Naming	22 (27)	20	19	17
Prototypes	19 (22)	18	18	9

2.2. Adult participants

We assessed colour category constancy in adults using the data for 5 German observers (3 women, 20–29 years old) in Olkkonen et al. (2010). We added measurements with 10 British adults (6 women, on average 22.5 years old) under neutral daylight to evaluate colour term maturity in British children. All adult participants were tested for red-green colour deficiencies using 12 Ishihara plates (Ishihara, 2004), to all of which they responded correctly.

2.3. Apparatus

Measurements were conducted in six nurseries (see Acknowledgements) and in the Sussex Colour Group lab at the University of Sussex. To obtain colourful illuminations, we put light filters in front of windows that were the only light source in the room (Fig. 1b). The filter was integrated in a wooden frame, which made it possible to transport and adapt them to the nurseries. LEE filters 35 *Light pink* and 138 *Pale green* were used to produce a reddish and greenish illumination, respectively (cf. Table 2). These were the same types of filters as those used by Olkkonen et al. (2010, p. 3), allowing us to compare our data from children with their adult data. We chose the red and green (rather than blue and yellow) illuminants to disentangle our experimental conditions from the variation of daylight between blue and yellow, cf. Fig. 1B of Olkkonen et al. (2010, p. 3) and our Fig. 2e.

Olkkonen et al. (2010, their Fig. 1B) had measured illuminant spectra as reflected by a reflectance standard before and after the behavioural measurements. Fig. 2a–c shows the illuminant spectra obtained through those measurements. When time constraints of nurseries allowed, we double-checked chromaticity and luminance measuring the light reflected by a white Munsell chip (N9.5), and, in later sessions, by a reflectance standard (Sphere Optics, Zenith Polymer Standard, 99%, 250-2500nm) using a colorimeter (Konica Minolta, CS-100A). The individual measurements are shown as red and black dots in Fig. 2d–f.

Table 2
Specifications of illumination. Chromaticity coordinates (x, y) as well as luminance (Y) correspond to the averages in Fig. 2e-f.

Illumination	filter	x	y	Y [cd/m^2]
Neutral	unfiltered daylight	0.3179	0.3411	246.1
Green	35 *Light pink*	0.2461	0.3591	247.6
Red	138 *Pale green*	0.3487	0.3119	212.5

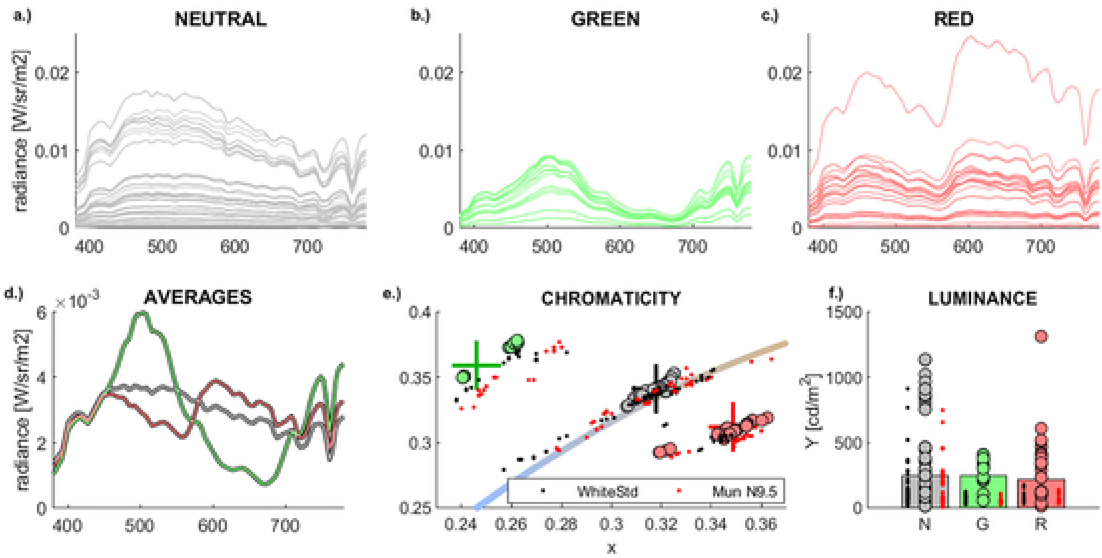

Fig. 2. Variation of illuminants within and across conditions. Panels a-c illustrate the illuminant spectra measured by Olkkonen et al. (2010) under neutral, green, and red illumination. Panel d shows the intensity-equated average spectra used for the simulations in Fig. 3. Panel e illustrates the variation of CIE1931 chromaticity coordinates across measurements; circles and large crosses correspond to the spectral measurements and their intensity-weighted averages (cf. panels a-c); small black and red dots refer to colorimetric measurements done before and after the sessions with children using a white standard or a white Munsell chip (N9.5); the blue-orange curve corresponds to the daylight locus. Panel f illustrates the variation of luminance across measurements. The bars show intensity-weighted averages.

Measurements with children were conducted between October and May, mostly during morning hours (8 h–13 h); for details see Fig. S2 in supplementary material. Daytime and seasonal variation and overcast, as well as the changing locations (nurseries) may explain the variation of illuminant luminance (Fig. 2f) and chromaticity along the daylight locus (blue-yellow curve in Fig. 2d-e; cf. Taylor & Kerr, 1941; Judd et al., 1964; Lee & Hernández-Andrés, 2005a, 2005b; Granzier & Valsecchi, 2014). Control of such variation was limited because sessions depended on nursery times and the schedule of children and nursery staff.

2.4. Stimuli

For measurements of categorisation in children, we made a subset of the maximally saturated Munsell chips used in previous studies on adult categorisation (such as Olkkonen et al., 2010; Berlin & Kay, 1969; originally proposed by Brown & Lenneberg, 1954; for review see Witzel, 2018a). This stimulus sample does not control for chroma and risks confounding effects of saturation and typicality (Witzel, Cinotti, & O'Regan, 2015; Witzel, 2018b; Witzel, Maule, & Franklin, 2019). Nevertheless, we opted for maximally saturated Munsell chips to optimise stimuli for mapping categories across hue and lightness and across illuminations in children because children are likely to produce more noisy categorisation data than adults. The chips varied across 40 Munsell hue steps and 5 lightness levels (Munsell Values 3–6, and 8), and were maximally colourful (Munsell chroma). Only 3 chips were included at lightness four to make sure the stimulus set contains unambiguous examples of red. The stimulus set thus consisted of overall 163 colours. Table 3 provides the specification of Munsell Chroma for each Munsell chip and highlights the stimulus set used with children in comparison with the full set used to obtain adult data.

The colour chips were set in cardboard chip holders in the shape of items of clothing. There were four sets of chip holders for all chips: trousers, hats, shirts, and boots. In each of the four sessions of repeated measurements, a different set of chip holders was used. This was done to prevent children associating the animal with the chip holder and thus the colour from an earlier session. For example, it could be confusing asking for blue trousers if a child remembers that the same animal wanted red trousers in a former session; but if the animal wants their hats, shirt, or boots in another colour in a later session, this would not contradict the red trousers in the previous session. We varied the pairing of chip holder sets with sessions and illumination conditions across child participants to minimise confounds.

All chip holders were painted in the grey of Munsell N5, and the chips in the chip holders were presented on a grey fabric as background to control for local contrast. The eight toy animals included a camel,

Table 3
Munsell Chroma for each chromatic chip. The numbers in the cells indicate Munsell Chroma; rows correspond to Munsell Value and columns to Munsell Hue. The 163 chips used with children are highlighted grey.

Hue	R				YR				Y				GY				G				BG				B				PB				P				RP			
Value		5		10		5		10		5		10		5		10		5		10		5		10		5		10		5		10		5		10		5		10
9	2	2	2	2	2	2	2	2	4	6	6	6	6	4	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
8	**6**	**6**	**6**	**6**	**6**	**6**	**8**	**14**	**16**	**14**	**12**	**12**	**12**	**10**	**10**	**8**	**8**	**6**	**6**	**6**	**6**	**4**	**4**	**4**	**4**	**4**	**4**	**4**	**6**	**6**	**4**	**4**	**4**	**4**	**6**	**6**	**6**	**6**	**6**	**6**
7	8	8	10	10	10	14	14	14	12	12	12	12	12	12	10	10	10	8	8	8	8	8	6	6	6	6	6	8	8	8	6	6	6	6	8	8	10	10	8	8
6	**12**	**12**	**12**	**14**	**16**	**12**	**12**	**12**	**10**	**10**	**10**	**10**	**10**	**10**	**12**	**12**	**10**	**10**	**10**	**10**	**8**	**8**	**8**	**8**	**8**	**8**	**8**	**10**	**10**	**10**	**8**	**8**	**8**	**8**	**10**	**10**	**10**	**10**	**12**	**12**
5	**14**	**14**	**14**	**16**	**14**	**12**	**10**	**10**	**8**	**8**	**8**	**8**	**8**	**8**	**10**	**12**	**12**	**10**	**10**	**10**	**10**	**8**	**8**	**8**	**8**	**8**	**8**	**10**	**12**	**12**	**10**	**10**	**10**	**10**	**10**	**12**	**12**	**12**	**14**	**14**
4	**14**	**14**	**14**	12	10	8	8	6	6	6	6	6	6	6	8	8	10	10	10	10	8	8	8	8	6	6	8	8	10	10	12	10	10	10	10	10	10	10	10	10
3	**10**	**10**	**12**	**10**	**8**	**6**	**6**	**6**	**4**	**4**	**4**	**4**	**4**	**4**	**6**	**6**	**8**	**8**	**8**	**8**	**6**	**6**	**6**	**6**	**6**	**6**	**6**	**8**	**10**	**10**	**12**	**10**	**10**	**10**	**10**	**10**	**10**	**10**	**10**	**10**
2	8	8	8	6	4	4	4	2	2	2	2	2	2	2	4	4	4	6	6	6	4	4	4	4	4	4	6	6	6	8	10	8	8	8	6	6	8	8	8	8
5	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6
5	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
5	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2

cat, dog, duck, monkey, pig, Mr. Squirrel, and Ms. Squirrel. Pairing of toy animals with colour terms was randomised.

For all measurements with adults, the same set of Munsell chips was used as in Olkkonen et al. (2010). This is the full set of 320 (40 hues, 8 lightness levels) maximally saturated Munsell chips, 10 achromatic chips, and a set of 120 desaturated chips (40 hues at Munsell values 5 and Munsell Chroma 2, 4, and 6).

To assess constancy of colour categorisation we computed the sensory colour signal reflected by each Munsell chip under each illumination. These colour signals are calculated based on the reflectance spectra of the Munsell chips, the spectra of the illuminations (*illuminants*), and the human cone sensitivities. First, the reflectance spectra for glossy Munsell chips were retrieved from the data base of the *Joensuu Color Group* (Kohonen, Parkkinen, & Jääskeläinen, 2006; Parkkinen, Hallikainen, & Jaaskelainen, 1989), which is now available via the University of Eastern Finland (https://sites.uef.fi/spectral/databases-software/spectral-database/). Second, we averaged the spectral measurements of Olkkonen et al. (2010; cf. Fig. 2a–c) to obtain illuminants for the computation. We equated the intensities of those spectra to yield the same overall radiance (area under the curve) for each of the three illuminants to control for the strong variation of intensity across measurements (Fig. 2d). Average CIE1931 chromaticities and luminance (xyY) are given in Table 2. Finally, we used the CIE1931 colour matching functions to calculate the colour signal (*Tristimulus Values*, XYZ), and represent them in CIELAB space. The resulting colours are illustrated by Fig. 3. Panels e and f illustrate the *illumination shifts*, i.e. how the colour signals is shifted across illuminations.

To make sure results do not depend on the choice of colour space, we double-checked results using CIELUV instead of CIELAB. The colours rendered with the spectra of the Finnish database show systematic changes in lightness and chroma across hue that do not exist in the actual Munsell chips; see light vertical bands across simulated chips in Fig. 3a-c (see also Derhak, Berns, & Derhak, 2012). For this reason, we also double-checked results using the Munsell renotation data (Newhall, Nickerson, & Judd, 1943) and calculating illumination shifts using the adaptation transforms in CIELAB and CIELUV-space rather than reflectance spectra. Results were very similar with the Finnish reflectance spectra and the renotation data in CIELAB and in CIELUV-space (for comparison see Fig. S6-S8). We focus on the results with reflectance spectra and CIELAB below.

2.5. Procedure

In the colour naming task, children assigned each of 163 Munsell chips to one of the 8 chromatic basic colour terms (pink, red, orange, yellow, green, blue, purple, and brown). Children were told that a toy animal would wear items of clothing only of one colour category (e.g., red). They were asked to give each animal their items of clothing. This was done in a two-step approach to make the task more accessible to children: First, the child was asked to find all the colours among those on the table that match a colour term. For this, the experimenter picked a random toy animal, and explained what colours the animal wants; for example: "this is the duck; the duck would like all the red trousers." The child could pick up the chips and drop them in a transparent plastic box. If the child hesitated with a chip in their hand, the experimenter repeated the above instructions (the duck would like…); when the child stopped picking any chips, the experimenter asked whether there were any "clothing items" corresponding to the respective colour term left among the chips, e.g. "are there any red trousers left for the duck?".

After having done this for each animal, the experimenter would pick up each of the remaining chips one by one, and ask the child which colour term it belongs to: "What colour is this one? Which animal would like this colour?" This was done until all chips were assigned to one of the colour terms. The time of the second phase varied depending on how many chips children spontaneously assigned to colour terms during the first phase.

After all chips were assigned to an animal, i.e., colour term, children identified the prototype of each category. For this, we spread all the colours of a category in front of the child. Then, we asked them to pick the colour chip that corresponds most strongly to the colour term, for example, the colour that is redder than all the other red colours. This was done for each colour term one after the other and in random order.

At any time during these tasks, children could correct their categories by moving chips from one animal/category to another until they were satisfied with the assignments to each animal/category. Some children also took breaks during the task, for example, because they wanted to show a toy in the nursery room to the experimenter. Yet, if a

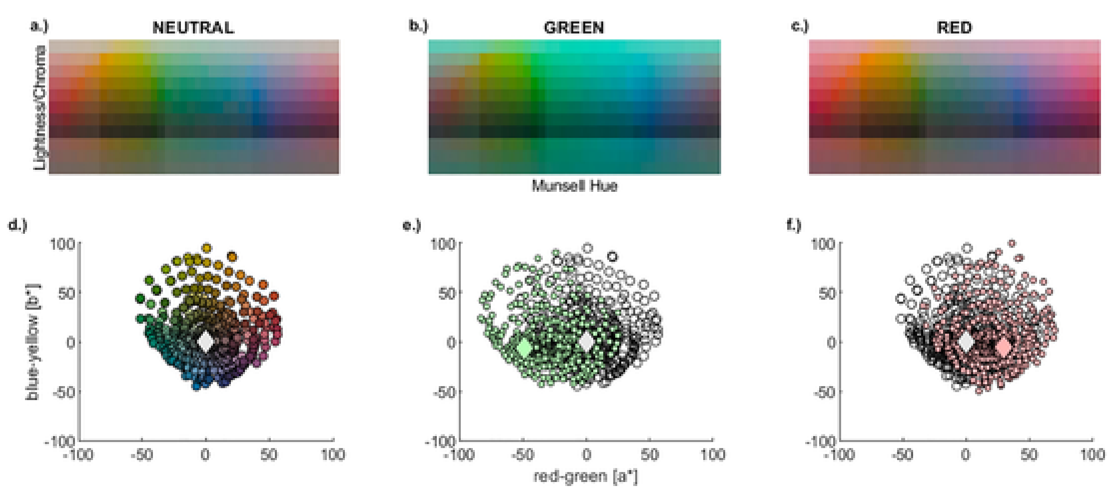

Fig. 3. Simulation of Munsell chips under the three different illuminations. The columns correspond to different illumination conditions, i.e. daylight without filter (a,d), green filtered (b,e) and red filtered daylight (c,f). The first row (a-c) illustrates the differences of colour signals corresponding to the Munsell chips under each illumination. Chips are ordered by their Munsell hue along the horizontal and by their Munsell Value (upper part at maximum Munsell Chroma) and Munsell Chroma (lower three rows at Munsell Value 5) along the vertical axis. Note that colours were rendered assuming an sRGB monitor and may not be accurate on other devices. Also note the vertical bands from yellow to green and from blue to pink (right side) that are an artifact of the spectral database (see main text). The second row represents the colour signals corresponding to the colours of the first row in the chromaticity plane of CIELAB space assuming the chromaticity of the neutral illuminant as the white-point. Each circle corresponds to one of the Munsell chips. The big diamonds identify the white-point. In the panels with chromatic illuminants (e,f), the colour signal under chromatic illumination and under the neutral illumination are shown by coloured (greenish in e or reddish in f) and by grey symbols, respectively. Note how colour signals are shifted under the chromatic compared to the neutral illumination (illumination shifts).

child took a break, they did not leave the room and hence kept their state of adaptation. Depending on a child's spontaneous categorisation and breaks, one session of measurements took between 30 and 60 min; two sessions took exceptionally long, namely 120 min (for details, see Fig. S2c).

Each session of measurements under one of the conditions was done on a different day. The two measurements under neutral illuminations (without filter) were done in the first and last sessions (i.e., days). The order of the measurements under the green and red illuminations changed randomly across children.

When chips were spread on the table they were presented roughly ordered by their hue (cf. Fig. 1a) to facilitate the task for children. We randomly changed between two spatial arrangements across measurements to avoid effects of chips at the left and right borders of the stimulus display. In the first arrangement, hues were ordered with red chips at the left and right borders, i.e., starting from 2.5R and ending with 10RP, as in the Munsell charts used in previous studies (see also Fig. 3a). In the second arrangement, the green–blue hues were at the borders, starting with 2.5 GB and ending with 10G (as in Fig. 1a).

Measurements in adults followed the protocol of Olkkonen et al. (2010): all chips were presented in a randomly mixed pile and observers were asked to sort them according to the basic colour terms. After that, they were asked to choose the chip that is the most typical example within each category. The five participants from Olkkonen et al. (2010) provided data once with a grey fabric as background (full local contrast condition) and once with a black fabric and wearing black gloves (reduced local contrast conditions) under each illumination condition. We pool these two conditions because they provided similar categorisation data (for details, see Olkkonen et al., 2010). Additional British adults were measured only once under neutral illumination (no filter) and with a grey cloth as background (full local contrast). For overview, see Table 1.

3. Results

Chromaticities are more greenish (reddish) under the green (red) illumination than under the neutral illumination (Fig. 3). Without colour constancy more chips would be named "green" under the green illumination, and more chips would be called "red" or "pink", and fewer "green", under the reddish illumination than under the neutral illumination. Fig. 4 illustrates aggregated colour categories in adults and children under neutral, green, and red illumination (a similar diagram for British adults under neutral illumination is provided in the supplementary Fig. S3). In Fig. 4, the green category seems slightly larger, and pink slightly smaller under green than under red illumination in both, adults (panels b-c) and children (panels e-f). Effects of the illumination seem to be stronger for the desaturated (lower rows C2-6) than for the other, maximally saturated chips. However, all analyses below focus on the 163 maximally saturated chips that were also measured with children (cf. Table 3).

We tested whether effects of illumination changes were reliable and whether they were correlated to colour term maturity. To quantify colour term maturity, we considered the consensus of colour term usage across adult observers as mature colour term usage. We determined whether an observer named a chip with the same colour term as the adult consensus. We did this for each chip and for each condition separately, and then averaged across conditions and chips to obtain a single maturity index for each observer (Witzel, Flack, et al., 2013). For British children, the adult consensus was determined as the mode across the 10 British adults (cf. Fig. S3). For comparison, we also calculated colour term maturity for each German adult comparing them to their mode colour naming. The higher this index the more the colour naming of an individual observer resembles the way most adults name colours. We doublechecked all Pearson correlations across participants with Spearman correlations to exclude spurious results due to outliers. Spearman correlations will only be reported if they do not support results with Pearson correlations.

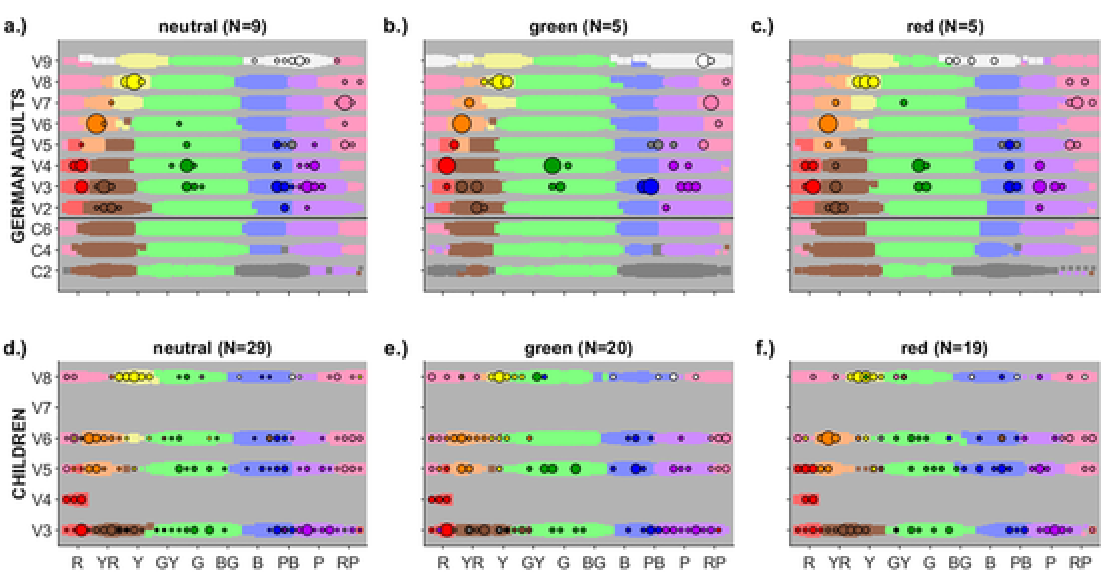

Fig. 4. Categorisation under different illuminations. Mode categories and their prototypes of German adults (first row) and (British) children (second row) separately under neutral (left column), green (centre column), and red (right column) illumination (For categories of British adults, see supplementary Fig. S3.). The areas in desaturated, light colours are composed of squares that correspond to the Munsell chips in Fig. 3a-c. Like there, the chips are ordered according to their hue along the x-axis and according to their lightness and chroma along the y-axis. Labels along the x-axis indicate the centres of principle and intermediate Munsell hues, labels along the y-axis refer to the level of Munsell Value (V) and Munsell Chroma (C). The colour and size of those squares indicate the mode colour names and the consistency of their use, respectively. The circles correspond to prototype choices, with their size representing the frequency of choosing that colour. In panels d-f, areas without any symbols indicate that those chips were not measured in children. The main analyses focus on those 163 chips that were measured in both, adults and children.

3.1. Naming consistency

As a first approach to quantify the effects of illuminations on colour naming we examined naming consistency across illumination changes. Consistency is calculated as $(k-1)/(n-1)$, where k is the frequency of mode colour terms and n the number of measurements. Hence, using different colour terms ($k = 1$) across all measurements results in a consistency of zero ($k - 1 = 0$). Consistency is first calculated for each chip, and then averaged across chips. Unlike maturity, the calculation of naming consistency is independent of the similarity to the adult consensus. Fig. 5 illustrates naming consistency for each child (circles) and adult (diamond) as a function of colour term maturity for different conditions.

To assess the effect of illumination changes, we compared naming consistencies in repeated measurements under the same illuminations with consistencies between the measurements under neutral and green, and between those under neutral and red illuminations. The aggregated data in Fig. 6a-b is based on the individual consistency data in Fig. 5. For adults, we also computed consistency across the repeated measurements (full and reduced local contrast) under the red and under the green illumination, respectively (dark green and dark red bars); for children such data was not available.

Naming consistencies were generally high in adults. Nevertheless, they were significantly smaller across illumination changes (neutral-red, neutral-green, green-red) than in repeated measurements under neutral illumination in paired t-tests (all $t(4) < 0.05$). Consistency of repeated measurements under green and red illuminations were only slightly smaller than under neutral illumination (not significant). In children, naming consistency was lower than in adults. This was shown through an independent *t*-test comparing consistency between children and adults in each of the four comparisons in Fig. 6b (all $p < 0.007$). Children's naming consistency was lower for the neutral-green ($t(15) = 3.9$, $p = 0.001$) and for the green–red ($t(12) = 2.2$, $p = 0.048$) than under neutral illumination. This difference was not significant for the neutral-red illumination change ($t(13) = 1.1$, $p = 0.28$). Note that the low number of degrees of freedom is due to the fact that not all children completed all three conditions (e.g. N1, N2, and R) necessary for each of those paired t-tests.

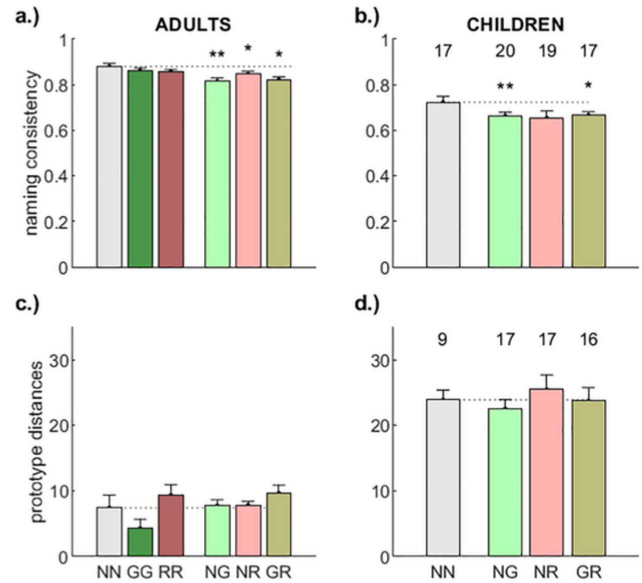

Fig. 6. Consistency of categorisation across illumination changes. The first row (a-b) shows naming consistencies in colour naming across repeated measurements under the same or under a different illumination. The second row represents distances between prototype choices across those repeated measurements. The left column (a,c) displays the results for the 5 German adults, the right column (b,d) those for children. Grey, dark green, and dark red bars correspond to comparisons between repeated measurements under neutral (NN), red (RR), and green (GG) illumination. Light green, red, and yellowish bars refer to comparisons between neutral and green (NR), neutral and red (NR), and green and red (GR) illuminations. Error bars indicate standard errors of mean. The dotted line corresponds to the repeated measurements under neutral illumination, and the asterisks indicate significant differences from those measurements in paired t-tests. * $p < 0.05$; ** $p < 0.01$. The numbers above the bars in panels b and d indicate the number of children that participated in both conditions involved in the calculation of consistency. Naming consistencies, but not prototype distances, are lower in repeated measurements across illuminations (right part of each diagram) than under the same illumination (left part). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

If children were less colour constant than adults, the effects of illumination change on naming consistencies should be larger for children than for adults. To test this idea, we compared the relative reduction of consistency due to illumination change between adults and children. To determine the relative reduction of consistency, we subtracted for each observer the consistency in repeated measurements under neutral illumination (grey bars in Fig. 6a-b) from the consistency across neutral and green, neutral and red, and across green and red illuminations (rightmost coloured bars in Fig. 6a-b). We calculated an independent *t*-test to compare children with adults. None of the tests showed a significant difference between adults and children (neutral-green: $t(19) = 0.05$, $p = 0.96$; neutral-red: $t(17) = 0.4$, $p = 0.68$; $t(16) = 0.69$, $p = 0.50$). So, there is no evidence from consistencies that children were less colour constant than adults.

As can be seen in Fig. 5, naming consistencies were correlated with colour term maturity under neutral illumination and under illumination changes (min $r(20) = 0.78$, all $p < 0.001$), indicating that observers are more consistent the more similar their naming is to the adult consensus. Slopes seem roughly similar under neutral illumination (Fig. 5a) and under illumination changes (Fig. 5b-c). It is thus difficult to in-

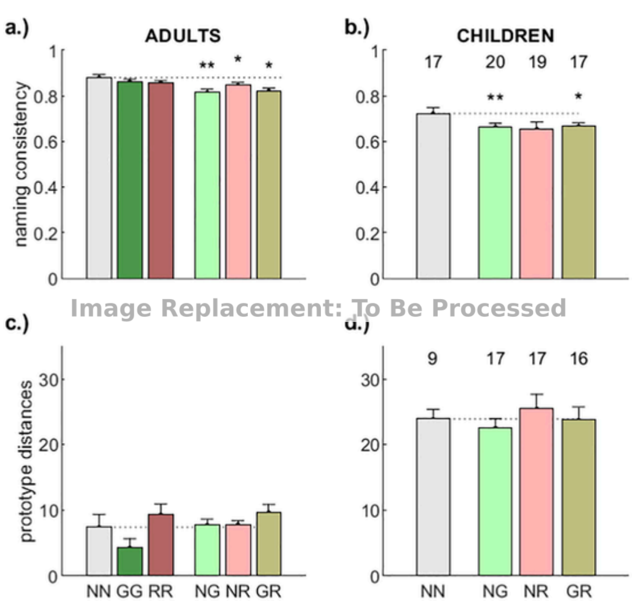

Fig. 5. Consistency and maturity across illuminations. Panels illustrate the relationship between colour term maturity (x-axis) and naming consistency (y-axis) in repeated measurements under neutral (a), across neutral and green (b), across neutral and red (c), and (d) across green and red illuminations. Each circle corresponds to a child, each diamond to a German adult. Regression lines and correlations (bottom left) are shown in the respective colour. The black regression line for the neutral illumination (a) is reproduced in panels b-c to allow comparing the slopes of regressions. *** $p < 0.001$. Consistency is strongly correlated with maturity in repeated measurements and across illumination changes. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

fer from those measures of naming consistency which developmental trend is specific to the illumination change, and which is not.

3.2. Constancy of category membership

For this reason, we separated the variation of categories that is specific to the change in illumination from general, unspecific inconsistencies in categorisation. We determined how the illumination change affected the colour signal corresponding to the colour chips and how that change in the colour signal would affect category membership if there was no colour constancy. For illustration of our approach, consider the example with aggregated adult categories in Fig. 7.

Besides the comparison between neutral and green, and neutral and red in Fig. 7, we also did these simulations for the comparison between green and red illumination (not shown in Fig. 7). We did these simulations for each individual observer. When there were multiple measurements for a condition (e.g. twice each condition in adults), we calculated each comparison (e.g. neutral1-red1, netural1-red2, netural2-red1, neutral2-red2), and averaged across the comparisons in each observer. For the green–red comparison, we calculated the changes from green to red, and the changes from red to green, and averaged across all of those comparisons.

If categorisation is affected by the illumination, shifted chips should be less consistent than stable chips. Fig. 8 compares the consistency of stable (light bars) and shifted chips (dark bars). In all three comparisons of illuminations, shifted chips were significantly less consistent than stable chips in paired t-tests for adults (min $t(4) = -3.7$, $p = 0.02$) and children (min $t(16) = -5.3$, $p < 0.001$).

Fig. 7a shows the aggregated categories under neutral illumination. Each Munsell chip corresponds to a specific sensory colour signal under neutral illumination. We simulated the colour signal of the chips under the green and red illuminants (see Method, Stimuli). We then determined, which colour signal under the neutral illumination (see Figure 3a,d) is closest (in CIELAB chromaticity) to each colour signal under the green (Figure 3b,e) and red illumination (Figure 3c,f). We assigned the category corresponding to the colour signal under the neutral illumination (Fig. 7a) to the chips with similar colour signals under the green and red illuminations. The resulting categories are shown in Fig. 7b-c for the simulation under the green and red illumination, respectively. Note that many chips are categorised as green and few as red under the green illumination because the green illumination makes the colour signal greener (Fig. 7c). The inverse is visible for the red illumination (Fig. 7b). Also note that not all colour signals are shifted strongly enough to change categories. In particular, chips at the centres of large categories, such as green, remain in the same category under both simulations; we will call these chips "stable" (see chips without black edge) and chips whose colour signal is shifted to another category "shifted". Only those "shifted" chips can be expected to change category membership due to the change in illumination.

Next, we wanted to test the question whether colour constancy increases with colour term maturity. Since consistency is correlated with maturity (Fig. 5), the question is whether there is any change in consistency that is specific to the illumination change. However, the consistency of shifted and stable chips was highly correlated for the green-neutral ($r(23) = 0.46$, $p = 0.02$), neutral-red ($r(22) = 0.92$, $p < 0.001$), and the green–red illumination change ($r(20) = 0.69$, $p = 0.0003$). To account for that, we calculated partial correlations between the colour term maturity and the consistency of shifted chips, while controlling for the consistency of stable chips (cf. Fig. 9a–c). There were positive correlations for the neutral-green ($r(23) = 0.62$, $p = 0.001$) and the neutral-red illumination change ($r(22) = 0.46$, $p = 0.03$); the correlation for the green–red illumination change was also positive, but missed significance ($r(20) = 0.37$, $p = 0.09$). Although these results seem to generally support the idea that colour category constancy increases with colour term maturity, they also reveal

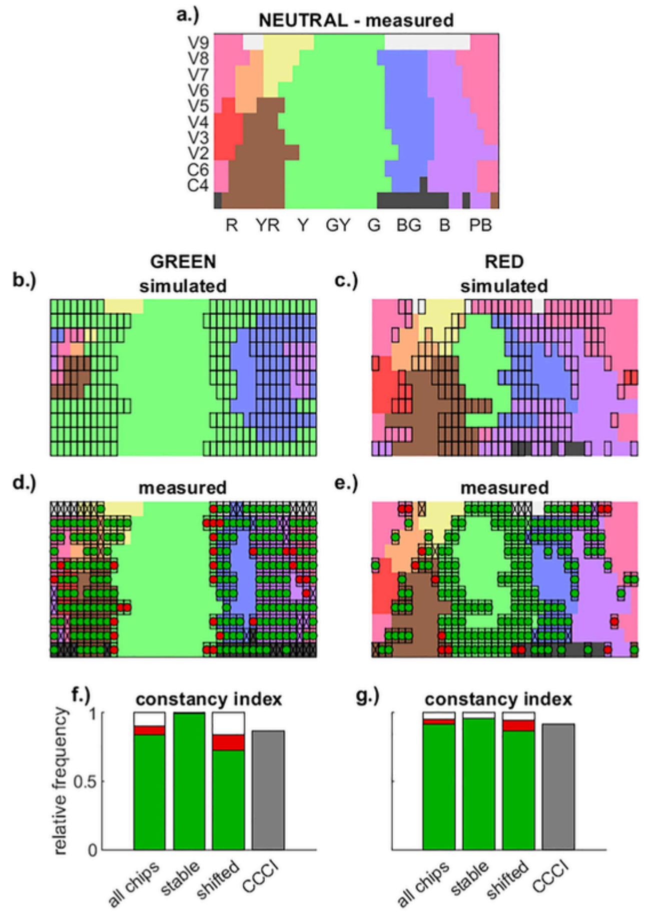

Fig. 7. Calculation of the colour category constancy index (CCCI). Panel a shows the aggregated colour categories of German adults under neutral illumination (as in Fig. 4.a). The second row (b-c) simulates colour categories in the complete absence of any constancy under the green (b) and red (c) illumination. Black edges highlight chips whose category membership differs compared to the neutral illumination (a) because they are shifted into other categories due to the effect of the illuminant on the colour signal. The third row (d-e) shows the categories produced by the German observers under the green and red illumination, respectively (as in Fig. 4.b-c). Green dots indicate whether the category membership of a chip is the same as the one under neutral illumination (panel a) and is hence colour constant. The red dots show whether category membership is the same as the simulated, non-constant categories (panel b and c). Black crosses indicate categories that neither correspond to the neutral categories in panel a, nor to the simulations in panels b and c. The last row (f-g) illustrates the calculation of the colour category constancy index (CCCI, grey bar). The proportions of the first three bars indicate how many chips are categorised under the green (f) and the red (g) illumination in the same way as under the neutral illumination in panel a (green part of the bar), as in the non-constancy simulation in panels b-c (red part), and in a way that disagrees with both, constancy and illumination shift (white part). The first bar ("all") refers to the proportions across all chips. The second and third bar separate between colour signals of chips that were and were not shifted ("shifted" vs "stable") into other categories due to the illumination change. The proportions indicated by the green, red, and white parts of the third bar correspond to the green and red circles and the crosses in the third row (d-e). The CCCI (grey bar) is the green proportion (constant chips) relative to the sum of the green and the red proportions (constant + non-constant) in the third bar. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

differences across illuminations: Under the neutral-red illumination change, consistency of stable chips almost completely controls for the maturity of adult individuals (black diamonds close to zero along the x-axis). In contrast, adult residuals follow the positive trend under the neutral-green and red-green illumination change. This suggests that

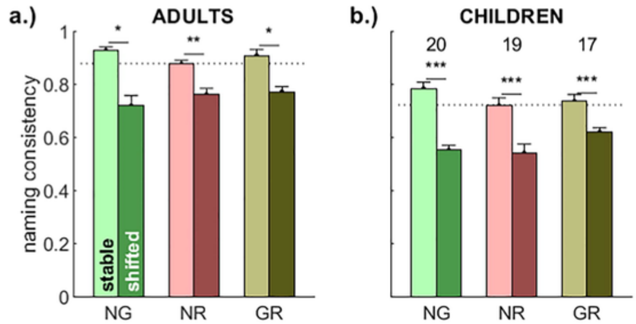

Fig. 8. Illumination-specific differences in consistency. The x axis refers to illumination conditions as in Fig. 6, the y-axis to average naming consistency. For further comparison, the consistency for repeated measurements is shown as a dotted line in the background (cf leftmost bar in Fig. 6a-b). Numbers above bars in panel b indicate the number of children that participated in both conditions involved in the calculation of consistency. In all comparisons, shifted colour chips (dark saturated bars) are named less consistently than stable chips (pale light bars).

consistency for stable chips influences the relationship between maturity and consistency of shifted chips in a complicated way.

To avoid these complications and to further isolate illumination-specific inconsistencies, we focused on the shifted chips alone. We evaluated whether the actual categorisation by the observers follows the category shift predicted by the illumination shift, or is the same as under neutral illumination. This is illustrated by Fig. 7d-e, which allows for comparing shifted chips (black edges) to categorisation (coloured areas) under green and red illumination, respectively. The measured categories of the observers may involve shifted chips in three different ways: first, observers may categorise shifted chips in the same way as under the neutral illumination (Fig. 7a); this would indicate that category membership is not affected by the illumination change and category membership of the respective chips is colour constant (see green dots in Fig. 7d-e); second, observers may categorise shifted chips in line with the illumination change, i.e. as predicted in Fig. 7b-c; this would contradict colour constancy (red dots in Fig. 7d-e). Finally, observers might also categorise them in a way that is different from categorisation under netural illumination (Fig. 7a) and from the effect of the illumination shift (Fig. 7b-c). The few chips that are categorised in such an unspecific way are uninformative about colour constancy (see crosses in Fig. 7d-e).

We created a colour category constancy index (CCCI) that takes only those chips into account that are specific to colour constancy, i.e. the

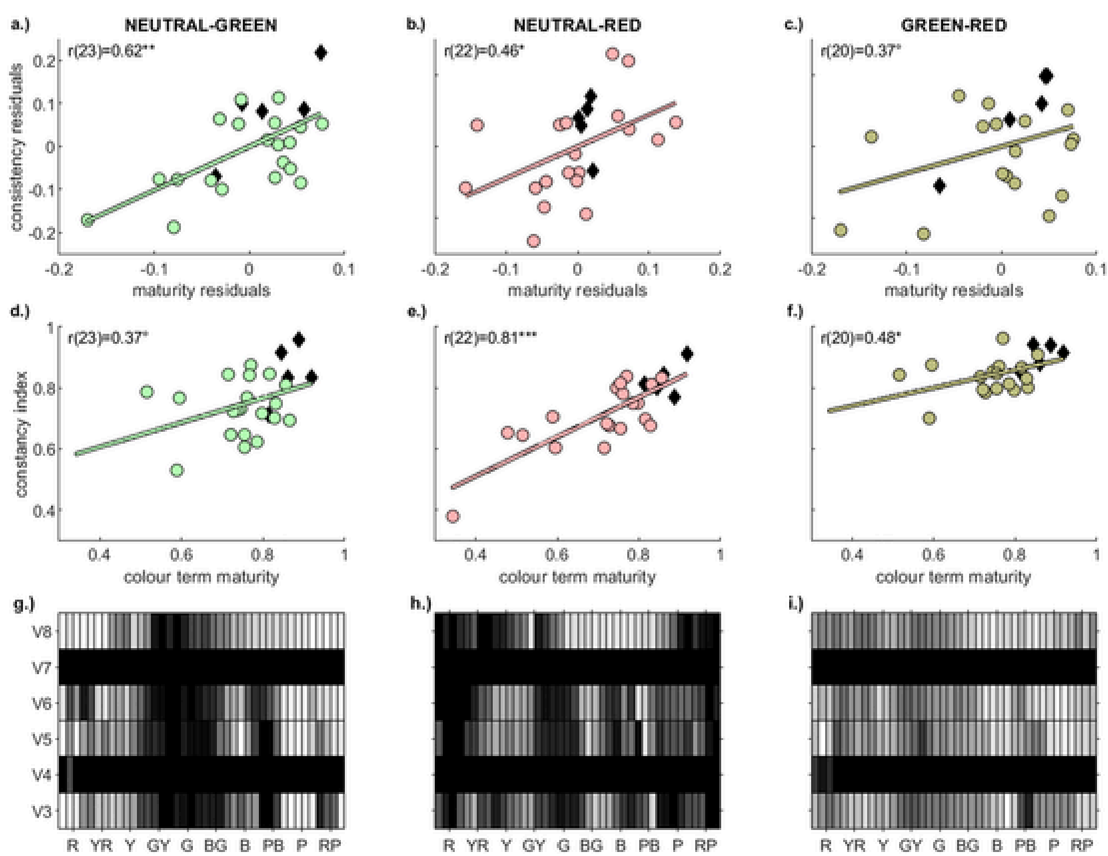

Fig. 9. The relationship between categorical colour constancy and colour term maturity. Columns refer to the comparison of colour naming between neutral and green illuminations (a, d, g), neutral and red illuminations (b, e, h), and green and red illuminations (c, f, i). The first row (a-c) shows, for each individual, naming consistency of "shifted" chips as a function of colour term maturity while controlling for the consistency of "stable" chips. The x- and the y-axis represent the residuals from the correlations with the consistencies of stable chips. The second row (d–f) displays the colour category constancy index (CCCI) as a function of colour term maturity. Circles show data of children, diamonds those of German adults for comparison. The regression line is shown and correlations between measures are reported in the upper left corner. Panels a–f suggest that there is a relationship between category colour constancy and colour term maturity. For the Discussion, the last row (g-i) illustrates the frequency, with which the colour signal of a chip is shifted into another category. Each rectangle corresponds to a Munsell chip, and its lightness indicates the relative frequency of becoming a "shifted chip" (cf. chips highlighted by black edges in Fig. 7b–e) when changing from neutral to green illumination (panel g), from neutral to red (h), and from green to red (i). These relative frequencies are calculated across all observers, i.e. adults and children together. The lighter the rectangle, the higher is the frequency; white indicates a frequency of one, black a frequency of zero or the absence of data. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

shifted chips that are either categorised as under neutral illumination or as predicted by the illumination shift. Fig. 7f-g illustrates the different types of proportions that can be identified in the categorisation data. The only proportions that are specific to colour constancy are the green and red parts of the shifted chips (third bar). The CCCI (grey bar in Fig. 7f-g) divides the number of constant chips (green part of that bar) by the number of all chips that are specific to colour constancy (the sum of the green and the red part).

We calculated the CCCI for the categorisation data of each observer and examined the relationship between colour term maturity and this colour constancy index (Fig. 9d–f). The CCCI was positively correlated with colour term maturity for the neutral-red ($r(22) = 0.78$, $p < 0.001$) and the green–red ($r(20) = 0.48$, $p = 0.02$) illumination shift; for neutral-green, it was also positive, but missed significance ($r(23) = 0.37$, $p = 0.06$). This result is surprising given the significance of partial correlations for neutral green, but not for red (Fig. 9a-b). In addition, the CCCIs of children were on average much higher for the green–red (Fig. 9.f) than for the neutral-green (Fig. 9.d) and neutral-red comparison (Fig. 9.e); paired t-tests showed that these differences were significant ($t(16) = 7.2$, $p < 0.001$; $r(16) = 5.3$, $p < 0.001$. This is also very much surprising because the chromaticity of the neutral illuminant is roughly in between the green and red ones (Fig. 2e), implying that the illumination change between green and red is largest (Figure 3e-f). We come back to this point in the Discussion.

3.3. Prototype constancy

To assess the constancy of category prototypes (*prototype constancy*), we measured the variation in prototype choices across illumination changes. For this, we represented prototypes in CIELAB colour space. Then we calculated for each observer and each category Euclidean distances between prototypes (*prototype distances*) across illumination changes and, for comparison, across repeated measurements under neutral illumination. Larger prototype distances imply lower colour stability of prototype choices.

There were no systematic differences in prototype choices across the eight chromatic categories (cf. Fig. S6 in the Supplementary Material). For this reason, we focus on the averages across the categories. Fig. 6c-d illustrates the average prototype distances in adults and children. If prototype distances were affected by illumination changes, they should be larger across illumination changes than across repeated measurements under neutral illumination. However, paired t-tests comparing the conditions with illumination change to the repeated measurements under neutral illumination were not significant (all $p > 0.66$). There was also no significant difference between children and adults in the effect of the illumination change according to independent t-tests (all $p > 0.47$). So, there is no evidence that both adult and child prototype choices varied more strongly due to the illumination changes.

Fig. 10 allows for appreciating whether the prototype choices of the children were shifted by the change in illumination. Each dot corresponds to the difference in chromaticity between prototype choices in different conditions. The coloured disks indicate the chromaticities of the illuminants under comparison. Fig. 10a refers to the repeated measurement under neutral illuminations, which involves only one illuminant chromaticity (grey disc in the centre). Fig. 10b-d show prototype choices across changes in illumination. The black line between the coloured disks illustrates the illumination shift of the white-point. If prototype choices were affected by the illumination change, the dots in Fig. 10b-d should be shifted along the illumination shift. This is not the case.

We also calculated Brunswick ratios to identify differences in prototype choices that are specific to the illumination shift (e.g. Foster, 2011; Weiss et al., 2017). See Figs. S6-S7 of the Supplementary Material for details. For children, the Brunswick ratio varied strongly between 0.5% and 1.5%, as would be expected from the unsystematic variation in Fig.

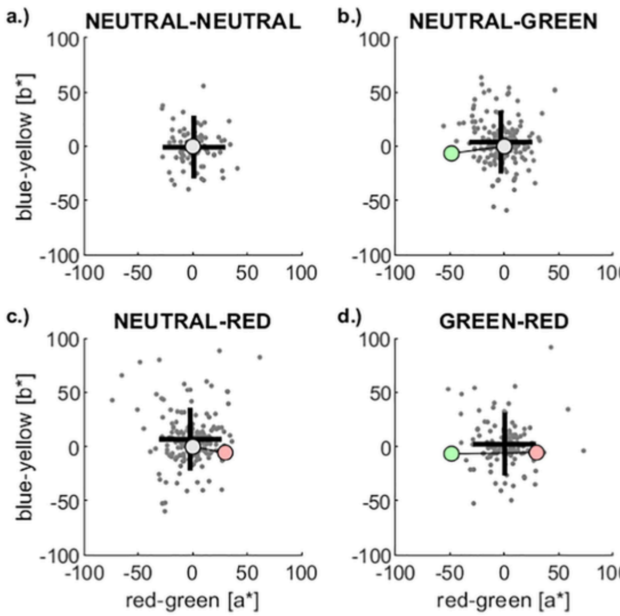

Fig. 10. Prototype differences in children and illumination shifts. The x- and y-axis represent the chromatic axes of CIELAB (cf. Fig. 3). The grey dots correspond to the difference in chromaticity between prototype choices of all categories and all children in the two measurements under neutral illumination (a), under green and neutral (b), red and neutral (c), and green and red illumination (d). Panels b and c also include comparisons with the second measurements under neutral illumination, hence the larger number of dots. The big black cross indicates the average difference. The coloured discs show chromaticities of the white-points, and the black line their illumination shift. Note that the prototype differences are randomly scattered around an average close to zero (neutral illuminant), rather than following the illumination shift. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

10. Taken together, the variation of prototype choices seemed to be random rather than being specific to the illumination changes.

4. Discussion

In sum, colour categories in children were generally stable across illumination changes (Fig. 4d–f). At the same time, naming in children was less consistent across than within illumination changes (Fig. 6). Non-constancy naming simulations (Fig. 7) showed that children varied strongly in colour category constancy at that age (Fig. 9). Some of the results suggested that colour constancy increased with colour term maturity (), but results were complicated because they differed between the green and red illumination (Fig. 9a-f).

4.1. Colour constancy in children

In adults, colour categories barely vary across illumination changes (Fig. 4a–c; Fig. 6a; see also, Olkkonen et al., 2010). Shifts in colour appearance due to adaptation and simultaneous contrast are known to strongly contribute to the constant appearance of surface colours across illuminations (Smithson, 2005; Foster, 2011; Witzel & Gegenfurtner, 2018b). The important role of adaptation is shown by the impressive aftereffects observers had when they left the experimental room after completing a session of about 1 h. Due to adaptation to the illumination in the experimental room, the neutral light in the hallway appeared in a saturated colour with the hue opponent to the one of the experimental illumination. Fig. 11 illustrates the effect using the photo of Fig. 1A in Olkkonen et al. (2010). The colour shifts are based on simulations in CIELAB assuming the white sheet of paper as an approximate reference white-point with the chromaticities for neutral and red illuminations in

Fig. 11. Illustration of adaptation. This figure illustrates the colour naming set-up for adult observers before (left) and after (right) adaptation (simulated in CIELAB). The average chromaticity of the sheet of paper in front of the monitor corresponds to the chromaticities of the red and neutral illuminant in Table 1 (rendered assuming sRGB); the white screen of the monitor illustrates how a white light looks when the observer is adapted to the red illumination. Images have been adapted from Fig. 1A in Olkkonen et al. (2010). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1. The monitor in the photo was set to monitor white and illustrates how a neutral white looks under the different states of adaptation: when observers just entered the room, they were not yet adapted to the red illumination, saw the white monitor as white, and the room as pink (left side in Fig. 11). After completing the experiment, they were adapted to the red illumination, and neutral white light, such as the one of the monitor, would appear green. When leaving the room, it would take a few of minutes for participants to readapt to neutral light, which causes the aftereffects.

In light of previous evidence, it seems likely that children at 2–4 years have already developed basic mechanisms of colour constancy, such as adaptation and simultaneous contrast induction (Dannemiller, 1989; Dannemiller & Hanko, 1987; Yang et al., 2013; Pereverzeva & Teller, 2009; Chien et al., 2006; Granrud et al., 2011). An informal observation during our measurements illustrates the strong effects of adaptation in one of the children (m3): at some point, he got distracted during the measurements under the green filtered light (cf. Fig. 1.b) and looked behind him at the door of the room. The boy was utterly surprised, pointed towards the doorsill and asked: "Why is it red under the door?" Of course, that light was just the common white light that leaked through the doorsill from the hallway of the nursery. It looked red because we were adapted to the complementary green illumination. The fact that the boy saw this light as red witnesses the effect of adaptation on his colour perception.

Consistent with the idea of adaptation and simultaneous contrast in children, our data showed high colour category constancy in children when compared to the variation of repeated measurements under the same illumination (Fig. 6). Nevertheless, there was a small but significant effect of the neutral-green and green–red illumination change on categorisation consistency.

We computationally isolated the effects of the illumination on colour categories through the colour category constancy index. This index disentangles illumination-specific effects on naming consistency from unspecific variation of consistency by focussing exclusively on colour chips and responses that are in line with the illumination change (Fig. 7). Except for one, all children reached colour category constancy above 50%, implying that their naming was robust to illumination changes for more than half of the respective chips (Fig. 9d–f).

In addition, we did not find any effect of the illumination on prototype choices. However, the lack of illumination-specific effects on prototype choices should not be interpreted as evidence for colour constancy. Compared to adults, children's prototype choices varied much more across repeated measurements independent of illumination changes. Adult average prototype distances were about 4.3–9.6 units in CIELUV space, which seems roughly comparable to adult discrimination thresholds for hue (e.g., Fig. 4 in Witzel & Gegenfurtner, 2018a) and chroma (e.g., Fig. 8 in Witzel et al., 2019). In contrast, average prototype distances of children were 22.5–25.5 (Fig. 6c-d). Single measurements of children varied up to 100 units, and variation was largely random (Fig. 10). Potential differences in colour discrimination between children and adults are too small to explain the large variation of prototype choices in children (Petzold & Sharpe, 1998; Ling & Dain, 2018). Instead, such large unsystematic variation might result if children did not understand the instructions, or if they did not yet develop a concept of typicality. Some of this variation might also be due to children being tired or bored at the end of the sessions when prototypes were measured. The resulting random noise might have covered the effect of the illumination change.

For these reasons, the prototype choices in children do not allow for conclusions about colour appearance and constancy. However, this does not undermine our results on categorisation, which indicate that 3–5 year-old children already possess a considerable degree of colour constancy.

4.2. Illumination-specific effects

In addition to the experimental manipulation with light filters, the chromaticity of the light shining on our stimuli varied along the daylight locus (Fig. 2e). Given the size of this variation, we cannot exclude that daylight variation affected children's consistency in colour categorisation. Previously, we had used a setup in the lab that fully controlled chromaticity and luminance of the illumination (Roger et al., 2020). However, this came with a trade-off for the naturalness and richness of the setting. We chose this setting here because it allows children to fully adapt to the illumination and to engage with the task in an environment they are used to, the nursery. The latter, presumably, helped maintaining children's motivation to categorise a large range of colour chips over a long period, sometimes up to 2 h (Fig. S2c). Also note, that the highest daylight variation in chromaticity and luminance occurred for the neutral condition without filter (cf. Fig. 2e-f); yet naming consistency is highest for the repeated measurements without filter (Fig. 6a). Most importantly, there were main effects of red and green filters on naming consistency despite the noise resulting from daily variation. Our results might have underestimated the categorisation differences between neutral, red, and green filters due to noisy daylight variation. If this is so, categorisation differences specific to the filter conditions might be still more pronounced than observed here, when measured in the absence of daylight variation.

Yet, the red and the green filters seemed to have different effects on naming consistency in our experiment. When computationally isolating the effects of the illumination on category shifts, we observed a strong relationship between colour constancy and colour term maturity for the neutral-red comparison; but this correlation was much lower for the green–red and missed significance for the neutral-green comparison. In addition, colour category constancy indices were higher for the green–red than for the other two conditions, although the effect of the illumination on the colour signal, the illumination shift, is largest for this condition.

With respect to these differences across conditions, we considered that the likelihood of a chip being shifted is related to how close it is to a category boundary where chips are also less consistent than at the category centre. To test for a possible relationship between consistency and illumination shifts, we looked more closely at the illumination shifts of different chips. Fig. 9g–i illustrates for each chip the relative frequency of becoming a "shifted chip" across all observers (i.e. adults and children). Under the neutral-green illumination change, chips are stable in the green category, but vary in the other regions (Fig. 9.g). These other regions comprise many small categories (e.g. red, orange, yellow, brown), implying that the frequency of shifted chips coincides with regions with comparatively many category boundaries. Since adult consistency is lower at category boundaries (cf. Fig. 4), the frequency of shifts for each chip is negatively correlated with the overall consistency in (German) adults ($r(161) = -0.42$, $p < 0.001$). In contrast, the neutral-red illumination shift affects green and blue chips with few category boundaries (Fig. 9.h), and there is no correlation between overall consistency and frequency of illumination shifts ($r(161) = -0.12$, $p = 0.14$). Under the green–red illumination shift, all chips seem to be shifted, but less so in the green category where naming consistency is high (Fig. 9.i). As a result, shift-frequency is also correlated with consistency under green–red ($r(161) = -0.31$, $p < 0.001$). These results show that the relationship between illumination shifts and consistency depends on the type of illumination. This may explain the different results for different types of illumination changes.

The correlation between illumination shifts and consistencies may potentially explain the differences across illumination comparisons in the correlations between maturity and measures of colour category constancy: In the neutral-green comparison, shifted chips are less consistent than stable chips because they coincide with boundaries (Fig. 8a and Fig. 9a); but since this consistency is not specific to the illumination shift, it disappears when focussing on illumination specific variation in shifted chips by using the colour category constancy index (Fig. 9d). In the green–red comparison, the illumination change affects the colour signals of much more chips, including those at the centre of categories (Fig. 9i). However, the centres of categories are particularly stable in naming (Fig. 4; see also, Fig. 8 in Olkkonen et al., 2010; Saji, Imai, & Asano, 2020; Witzel, Flack, et al., 2013). This could explain why the number of constant chips among the "shifted" chips is particularly high in the green–red comparison despite the large difference in illumination colour.

However, other studies that did not measure constancy through colour categorisation, also found differences across illuminate colours (Delahunt & Brainard, 2004; Daugirdiene, Kulikowski, Murray, & Kelly, 2016; Weiss et al., 2017). A difference across illumination colours was also found in a recent study on colour constancy in children (Wedge-Roberts et al., 2020). Hence, the intricate relationship between illuminant spectra, colour constancy, and category shifts requires further investigation.

4.3. Colour constancy and colour term acquisition

Admittedly, our evidence for a relationship between colour constancy and colour term acquisition is complicated by the differences between green and red illumination. If the consistency of "shifted" chips is affected by the location of category boundaries, the observed correlations between colour term maturity and colour category constancy (Fig. 9a–f) may be due to the sharpening of category boundaries during development (Raskin, Maital, & Bornstein, 1983; Saji et al., 2020; Witzel, Flack, et al., 2013). Since illumination induced shifts of colour signals mainly affect adjacent categories (Fig. 9g–i), such shifts may be confused with the uncertainty of category membership at the boundary. Sharpening during colour term acquisition implies that fewer chips are confused between adjacent categories. This, rather than the development of colour constancy, could be the reason for higher consistency of shifted chips with higher colour term maturity. In addition, prototype choices varied massively in children, speaking against the idea that prototypes are anchors of colour categories across illumination changes.

Hence, the present study hinges upon the success of our computational approach to separate colour constancy from unspecific inconsistencies in colour naming: if our colour constancy measure does not fully satisfy its purpose and is instead affected by unspecific variation, the correlation with colour maturity would not be specific to colour constancy. Yet, our observation of a relationship between colour constancy and colour term maturity complements prior findings of our research team (Rogers et al., 2020). In that study, we used a completely different method and found a relationship between colour constancy and colour naming when measuring each through a separate task. While that study was limited to a small range of colours under only one illumination change, the present study involves a very large range of colour chips and three illumination changes. Together, these two studies provide strong support to the idea that colour constancy and colour term acquisition are related.

The relationship between colour term acquisition and colour constancy supports the idea that colour term acquisition involves learning to handle multivalent information, which is also the basis of inferential colour constancy. While the mechanisms underlying constant appearance are likely to be part of early sensory mechanisms, such as adaptation, inferential constancy requires understanding how surface colours change across illuminations based on the experience of such colour changes (Witzel & Gegenfurtner, 2018b; Witzel et al., 2016). For this, a child needs to learn which surface colours seen under different illuminations are the same even if they appear differently. The processes of learning and calibration behind colour term acquisition and colour constancy may mutually benefit each other: on the one hand, improvements in colour constancy reduce uncertainty about surface colours and hence support calibrating the category membership of those colours; on the other hand, colour naming may support colour constancy by identifying surface colours across illuminations through communication.

5. Conclusion

We investigated whether children at the age of colour term acquisition are colour constant, and whether the development of colour naming is related to colour constancy. Similar to adults, categorisation in children was barely affected by strong changes in illumination. Yet, the effects of the illumination on colour categorisation scaled with colour term maturity, suggesting a relationship between colour constancy and colour term acquisition. These findings suggest that children progressively fine-tune and recalibrate the meaning of colour terms through processes of coordination and adaptation that may also support colour constancy. The relationship between colour term acquisition and colour constancy supports the idea that colour term acquisition involves learning to handle multivalent information (a colour term may refer to many different colours), which is also the basis of inferential colour constancy (surface colours can look different depending on the illumination).

CRediT authorship contribution statement

Christoph Witzel: Conceptualization, Funding acquisition, Investigation, Methodology, Data curation, Formal analysis, Project administration, Supervision, Visualization. **Zoe Flack:** Investigation, Writing - review & editing. **Emma Sanchez-Walker:** Investigation, Writing - review & editing. **Anna Franklin:** Funding acquisition, Project administration, Resources, Supervision, Writing - review & editing.

Acknowledgements

We are grateful to the nurseries Abracadabra and Chappell Croft in Worthing, Annan School in Uckfield, Busy Bees in Lewes, Little Tinkers in Mayfield, and St Aubyns School in Rottingdean for their cooperation. We thank Oshrat Meshulam-Madders and Ying Chen for help with data collection. This research was supported by a DAAD postdoc fellowship to CW (ref D/11/45774), and ERC funded project "CATEGORIES" to AF (ref 283605).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.visres.2021.05.008.

References

Berlin, B., & Kay, P. (1969). *Basic color terms: Their universality and evolution*. Berkeley, CA: University of California Press.
Brown, R. W., & Lenneberg, E. H. (1954). *A study in language and cognition. Journal of Abnormal and Social Psychology*, *49*(3), 454–462.
Chien, S. H., Bronson-Castain, K., Palmer, J., & Teller, D. Y. (2006). *Lightness constancy in 4-month-old infants. Vision Research*, *46*(13), 2139–2148. https://doi.org/10.1016/j.visres.2005.11.027. doi: S0042-6989(05)00668-1 [pii].
Cohen, J. B., & Kappauf, W. E. (1982). *Metameric color stimuli, fundamental metamers, and Wyszecki's metameric blacks. American Journal of Psychology*, *95*(4), 537–564.
Dannemiller, J. L. (1989). *A test of color constancy in 9- and 20-week-old human infants following simulated illuminant changes. Developmental Psychology*, *25*(2), 171–184.
Dannemiller, J. L., & Hanko, S. A. (1987). *A test of color constancy in 4-month-old human infants. Journal of Experimental Child Psychology*, *44*(2), 255–267. doi: 0022-0965(87)90033-6 [pii].
Daugirdiene, A., Kulikowski, J. J., Murray, I. J., & Kelly, J. M. F. (2016). *Test illuminant location with respect to the Planckian locus affects chromaticity shifts of real Munsell chips. Journal of the Optical Society of America A*, *33*(3), A77–A84. https://doi.org/10.1364/JOSAA.33.000A77.
Delahunt, P. B., & Brainard, D. H. (2004). *Does human color constancy incorporate the statistical regularity of natural daylight?. Journal of Vision*, *4*(2), 57–81. 10:1167/4.2.1/4/2/1/[pii].
Derhak, M. W., Berns, R. S., & Derhak, M. W. (2012). *Analysis and correction of the Joensuu Munsell Glossy spectral database. Color and Imaging Conference*, *2012*(1), 191–194.
Douven, I., Wenmackers, S., Jraissati, Y., & Decock, L. (2016). *Measuring graded membership: The case of color. Cognitive Science*, *41*(3), 686–722. https://doi.org/10.1111/cogs.12359.
Foster, D. H. (2011). *Color constancy. Vision Research*. https://doi.org/10.1016/j.visres.2010.09.006. doi: S0042-6989(10)00440-2 [pii].
Gelman, S. A., & Meyer, M. (2011). *Child categorization. Wiley Interdisciplinary Reviews: Cognitive Science*, *2*(1), 95–105.
Granrud, C., Corrow, S., Gilchrest, A., Mathison, J., & Yonas, A. (2011). *Four-month-old infants exhibit lightness constancy. Journal of Vision*, *11*(11). 416.
Granzier, J. J. M., & Valsecchi, M. (2014). *Variations in daylight as a contextual cue for estimating season, time of day, and weather conditions. Journal of vision*, *14*(1). https://doi.org/10.1167/14.1.22.
Hurlbert, A. C. (2007). *Colour constancy. Current Biology*, *17*(21), R906–R907. https://doi.org/10.1016/j.cub.2007.08.022. doi: S0960-9822(07)01839-8 [pii].
Ishihara, S. (2004). *Ishihara's Tests for Colour Deficiency*. Tokyo, Japan: Kanehara Trading Inc.
Judd, D. B., Macadam, D. L., Wyszecki, G., Budde, H. W., Condit, H. R., Henderson, S. T., & Simonds, J. L. (1964). *Spectral distribution of typical daylight as a function of correlated color temperature. Journal of the Optical Society of America*, *54*(8), 1031–1036.
Kohonen, O., Parkkinen, J., & Jääskeläinen, T. (2006). *Databases for spectral color science. Color Research and Application*, *31*(5), 381–390. https://doi.org/10.1002/col.20244.
Lee, R. L., & Hernández-Andrés, J. (2005a). *Colors of the daytime overcast sky. Applied Optics*, *44*(27), 5712–5722.
Lee, R. L., & Hernández-Andrés, J. (2005b). *Short-term variability of overcast brightness. Applied Optics*, *44*(27), 5704–5711.
Lindsey, D. T., & Brown, A. M. (2014). *The color lexicon of American English. Journal of vision*, *14*(2). https://doi.org/10.1167/14.2.17.
Ling, B. Y., & Dain, S. J. (2018). *Development of color vision discrimination during childhood: Differences between Blue–Yellow, Red–Green, and achromatic thresholds. Journal of the Optical Society of America A*, *35*(4), B35–B42. https://doi.org/10.1364/JOSAA.35.000B35.
Logvinenko, A. D., Funt, B., Mirzaei, H., Tokunaga, R., & Osorio, D. (2015). *Rethinking Colour Constancy* e0135029. *PLoS ONE*, *10*(9). https://doi.org/10.1371/journal.pone.0135029.
Newhall, S. M., Nickerson, D., & Judd, D. B. (1943). *Final Report of the O.S.A. Subcommittee on the Spacing of the Munsell Colors. Journal of the Optical Society of America*, *33*(7), 385–418.
Olkkonen, M., Hansen, T., & Gegenfurtner, K. R. (2009). *Categorical color constancy for simulated surfaces. Journal of Vision*, *9*(12). 6 6.
Olkkonen, M., Witzel, C., Hansen, T., & Gegenfurtner, K. R. (2010). *Categorical color constancy for real surfaces. Journal of Vision*, *10*(9), 1–22. https://doi.org/10.1167/10.9.16. doi: 10.9.16 [pii].
Owen, K., & Barnes, C. (2019). *The development of categorization in early childhood: A review. Early Child Development and Care*, *191*(1), 13–20. https://doi.org/10.1080/03004430.2019.1608193.
Parkkinen, J. P. S., Hallikainen, J., & Jaaskelainen, T. (1989). *Characteristic spectra of Munsell Colors. Journal of the Optical Society of America a-Optics Image Science and Vision*, *6*(2), 318–322. https://doi.org/10.1364/Josaa.6.000318.
Pereverzeva, M., & Teller, D. Y. (2009). *Simultaneous color contrast in 4-month-old infants. Perception*, *38*(1), 30–43.
Petzold, A., & Sharpe, L. T. (1998). *Hue memory and discrimination in young children. Vision Research*, *38*(23), 3759–3772. doi: S0042-6989(98)00047-9 [pii].
Philipona, D. L., & O'Regan, J. K. (2006). *Color naming, unique hues, and hue cancellation predicted from singularities in reflection properties. Visual Neuroscience*, *23*(3–4), 331–339. https://doi.org/10.1017/S0952523806233182. doi: S0952523806233182 [pii].
Pitchford, N. J., & Mullen, K. T. (2002). *Is the acquisition of basic-colour terms in young children constrained?. Perception*, *31*(11), 1349–1370.
Raskin, L. A., Maital, S., & Bornstein, M. H. (1983). *Perceptual categorization of color: A life-span study. Psychological Research Psychologische Forschung*, *45*(2), 135–145.
Rogers, M. R., Witzel, C., Rhodes, P., & Franklin, A. (2020). *Color constancy and color term knowledge are positively related during early childhood. Journal of Experimental Child Psychology*, *196*, 104825. https://doi.org/10.1016/j.jecp.2020.104825.
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), e12813. https://doi.org/10.1111/cogs.12813.
Smithson, H. E. (2005). *Sensory, computational and cognitive components of human colour constancy. Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *360*(1458), 1329–1346. https://doi.org/10.1098/rstb.2005.1633. doi: PX26MA7W586VQ2A7 [pii].
Steels, L., & Belpaeme, T. (2005). *Coordinating perceptually grounded categories through language: A case study for colour. Behavioral and Brain Sciences*, *28*(4), 469–489.
Taylor, A. H., & Kerr, G. P. (1941). *The distribution of energy in the visible spectrum of daylight. Journal of the Optical Society of America*, *31*(1), 3–8. https://doi.org/10.1364/JOSA.31.000003.
Vazquez-Corral, J., O'Regan, J. K., Vanrell, M., & Finlayson, G. D. (2012). *A new spectrally sharpened sensor basis to predict color naming, unique hues, and hue cancellation. [Research Support, Non-U.S. Gov't]* 7. *Journal of Vision*, *12*(6). https://doi.org/10.1167/12.6.7.
Wedge-Roberts, R., Aston, S., Kentridge, R., Beierholm, U., Nardini, M., & Olkkonen, M. (2020). *Developmental changes in colour constancy and the use of daylight illumination priors* 1106 1106. *Journal of Vision*, *20*(11). https://doi.org/10.1167/jov.20.11.1106.
Weiss, D., Witzel, C., & Gegenfurtner, K. (2017). *Determinants of colour constancy and the Blue Bias. i-Perception*, *8*(6). https://doi.org/10.1177/2041669517739635. 2041669517739635.
Witzel, C. (2018a). *Misconceptions about colour categories. Review of Philosophy and Psychology*, *10*(3), 499–540. https://doi.org/10.1007/s13164-018-0404-5.
Witzel, C. (2018b). *The role of saturation in colour naming and colour appearance*. In L. W., MacDonald, C. P., Biggam, & G. V., Paramei (Eds.), *Progress in Colour Studies: Cognition, language and beyond* (pp. 41–58). Amsterdam/Philadelphia: John Benjamin Publishing Company.
Witzel, C., Cinotti, F., & O'Regan, J. K. (2015). *What determines the relationship between color naming, unique hues, and sensory singularities: Illuminations, surfaces, or photoreceptors?. Journal of Vision*, *15*(8), 19. https://doi.org/10.1167/15.8.19.
Witzel, C., Flack, Z., & Franklin, A. (2013, 8-12. July 2013). Categorical colour constancy during colour term acquisition. In Paper presented at the AIC2013 - 12th international AIC congress, Newcastle upon Tyne, UK.
Witzel, C., & Gegenfurtner, K. R. (2018a). *Are red, yellow, green, and blue perceptual categories?. Vision Research*, *151*, 152–163. https://doi.org/10.1016/j.visres.2018.04.002.
Witzel, C., & Gegenfurtner, K. R. (2018b). *Color perception: Objects, constancy, and categories. Annual Review of Vision Science*, *4*(1), 475–499. https://doi.org/10.1146/annurev-vision-091517-034231.
Witzel, C., Maule, J., & Franklin, A. (2019). *Red, yellow, green, and blue are not particularly colorful* 27. *Journal of Vision*, *19*(14). https://doi.org/10.1167/19.14.27.
Witzel, C., Sanchez-Walker, E., & Franklin, A. (2013). *The development of categorical colour constancy. Perception*, *42*(ECVP Abstract Supplement), 19.
Witzel, C., van Alphen, C., Godau, C., & O'Regan, J. K. (2016). *Uncertainty of sensory signal explains variation of color constancy. Journal of Vision*, *16*(15), 8. https://doi.org/10.1167/16.15.8.
Yang, J., Kanazawa, S., Yamaguchi, M. K., & Kuriki, I. (2013). *Investigation of color constancy in 4.5-month-old infants under a strict control of luminance contrast for individual participants. Journal of Experimental Child Psychology*, *115*(1), 126–136. https://doi.org/10.1016/j.jecp.2012.11.013.
Yang, J., Kanazawa, S., Yamaguchi, M. K., & Motoyoshi, I. (2015). *Pre-constancy vision in infants. Current Biology*, *25*(24), 3209–3212. https://doi.org/10.1016/j.cub.2015.10.053.