Linguistic and Cultural Variation in Early Color Word Learning

Samuel H. Forbes [ID] and Kim Plunkett
University of Oxford

When and how do infants learn color words? It is generally supposed that color words are learned late and with a great deal of difficulty. By examining infant language surveys in British English and 11 other languages, this study shows that color word learning occurs earlier than has been previously suggested and that the order of acquisition of color words is similar in related languages. This study also demonstrates that frequency and syllabic complexity can be used to predict variability in infant color word learning across languages. In light of recent evidence indicating that color categories have universal biological foundations, these findings suggest that infants' experience and linguistic exposure drive their shift to culturally and linguistically mediated adult-like understandings of color words.

The domain of color perception and categorization has played a central role in furthering our understanding of the impact of language on cognition, and of cognition on language for over 60 years (Berlin & Kay, 1969; Brown & Lenneberg, 1954; Cohen, Chaput, & Cashon, 2002; Heider, 1972; Roberson, Davidoff, Davies, & Shapiro, 2005; Roberson, Davies, & Davidoff, 2000). In normal discourse, adult speakers treat colors categorically by grouping them into blocks linguistically, despite their continuous nature. With mounting evidence of emergent color categories in infants (e.g., Franklin, Drivonikou, Bevis, et al., 2008; Skelton, Catchpole, Abbott, Bosten, & Franklin, 2017) that may adapt as color terms are learned (Franklin, Drivonikou, Clifford, et al., 2008), it appears color categories have a strong biological component. However, questions still arise as to their formation. What brings about the change from infant biological color categories to adult categories? What contribution remains for linguistic and cultural components in setting boundary conditions on the learning process? Does developing from a universal color category imply a universal order in the learning of color categories?

Explanations of the formation of color categories have been the source of much debate. On the one hand, evidence from cross-linguistic differences in the perception of color have shown how categorical perception (CP) can differ by language group (Roberson, Hanley, & Pak, 2009; Roberson, Pak, & Hanley, 2008), giving weight to the idea that color categories are formed culturally (Roberson et al., 2000). On the other, analyses of World Color Survey data (Kay, Berlin, Maffi, Merrifield, & Cook, 2011) have shown universal similarities in color naming across different languages (Abbott, Griffiths, & Regier, 2016; Kay, 2003; Regier, Kay, & Cook, 2005; Regier, Kay, & Khetarpal, 2007).

Evidence from infant experiments have provided a fresh perspective on this debate. Infants have been found to possess CP of color in the right hemisphere (Franklin, Drivonikou, Bevis, et al., 2008), as opposed to the left hemisphere in adults (Gilbert, Regier, Kay, & Ivry, 2006). Similar results were found comparing toddlers who had not learned color terms, and therefore behaved like infants, with those who had, thus behaving like adults (Franklin, Drivonikou, Clifford, et al., 2008).

Recently, strong evidence has been reported for biological, prelinguistic color categories in a novelty-preference task, suggesting the presence of infant color categories (Skelton et al., 2017). Skelton

et al. found that when the results were plotted in a color space representative of the retinogeniculate pathways that makeup color vision, most of the categorical distinctions made by infants were separated by the axes in that color space. The results of that study therefore suggested that there is a strong association between the cardinal mechanisms of color vision and the way in which infants categorize colors. The infant color categories were found to be similar to the category centroids of nonindustrialized languages, suggesting some commonality with adult color categories. However, there is great diversity in how the color spectrum is divided across languages, suggesting that at some point, language and culture intervene and change the way color is categorized from the original, biological infant categories, to make it more relevant to the language and culture in question. The diversity between languages in the number of color words used may also lead to variability in the timing of acquisition of these terms.

The fact that there is a necessary transition from infant color categories to adult color categories (Skelton et al., 2017) may also provide further insights into why color word learning is perceived to be difficult (Franklin, 2006; Johnson & Huettig, 2011; Mervis, Bertrand, & Pani, 1995; Pitchford & Mullen, 2003; Soja, 1994; Wagner, Dobkins, & Barner, 2013). Researchers have argued that infants learn color words relatively late compared to other classes of words (Heider, 1971; Shatz, Behrend, Gelman, & Ebeling, 1996; Soja, 1994) and that early color word usage is riddled with haphazard, random usage (Pitchford & Mullen, 2003; Sandhofer & Smith, 1999). Explanations have focused on the need to learn to categorize the continuous spectrum of color (Kowalski & Zimiles, 2006) or the dominance of shape over color as a salient dimension (Sandhofer & Smith, 1999). Despite this speculation, the reported age of acquisition of color terms seems to have dropped dramatically in recent decades (Franklin, 2006; Shatz et al., 1996). That color word learning really is difficult is thus worthy of reassessment.

The vast majority of studies completed to date were behavioral assessments of color word knowledge. Many of the previous enquiries into toddler color word learning relied on color matching tasks, where participants were either asked to find an object matching a given color label or asked to match two objects by color (e.g., Sandhofer & Smith, 1999; Shatz et al., 1996; Soja, 1994). For toddlers as young as 2–3 years of age, this requires concentration and cooperation beyond simply comprehending color categories and color terms, and

also requires them to feel comfortable in the experimental setting. The findings of these enquiries found that toddlers could comprehend basic color terms usually by around 30 months of age, or slightly younger. Other past studies have examined toddlers' ability to produce basic color terms (e.g., Pitchford & Mullen, 2002, 2003), finding that their ability to produce some color terms was in place at 2–3 years of age but still improving up until 4–5 years. Producing terms in front of an experimenter could also be daunting for a toddler, giving these tasks an additional degree of difficulty. Thus assessments made by caregivers, with whom the toddlers are most comfortable, may be a comparatively sensitive measurement, reflected by the fact that diary studies have found color word comprehension as early as just over 2 years of age (Mervis et al., 1995).

The claim that color categories have a biological root in infancy (Skelton et al., 2017), as well as the claim that color categories in adults possess a common root in infant color categories might suggest a universal order of color term learning based on this biological root (see Bornstein, 1985; O'Hanlon & Roberson, 2006; Pitchford & Mullen, 2002). Although Berlin and Kay (1969) found a general order in which color words were developed by languages, no substantive evidence has been found to suggest that this might be mirrored in the way that infants learn color categories. Evidence for a systematic, universal order of color word learning has been equivocal (e.g., Andrick & Tager-Flusberg, 1986; Pitchford & Mullen, 2002; Shatz et al., 1996), suggesting that there have been some trends observed but no overarching pattern. Most of these studies did not consider a wide range of languages, thus also limiting the ability to test whether there are emergent patterns in color word learning.

In light of these findings, this study explores some linguistic and cultural determinants for learning early color words and color categories in different languages, and analyses the developmental profile in color word learning in order to examine the presence or absence of a universal order of color word learning. To address these issues, data were employed from existing parental surveys of children's word learning, also known as Communicative Development Inventories (CDIs; Fenson et al., 1994). CDIs are generally considered valid and reliable indicators of infant word learning (Bates, Bretherton, & Snyder, 1988; Dale, 1991; Dale, Bates, Reznick, & Morisset, 1989; Fenson et al., 1994; Mills, Coffey-Corina, & Neville, 1993, 1997; Styles & Plunkett, 2009). CDIs have the advantage

of being able to measure infant vocabulary on a large scale, with parents often asked to assess one or both of comprehension and production for each word. There has been considerable debate about the reliability of this measurement, however, particularly when measuring comprehension, especially for terms as abstract as color words (Houston-Price, Mather, & Sakkalou, 2007; Tomasello & Mervis, 1994). Studies have consistently found that CDIs measure word learning quite consistently when compared with normative developmental scales (Dale, 1991; Dale et al., 1989), and have been confirmed with measurement from ERPs (Mills et al., 1993, 1997). More recently, CDI validity of comprehension was tested against behavioral measurement (Styles & Plunkett, 2009), finding that parents are quite conservative in their determinations of whether the infant comprehends a term and that comprehension, as measured by parental report, is quite an accurate measurement.

In Study 1, we investigate toddler's comprehension and production of basic color words in British English. If indeed color word comprehension learning occurs as late as over 2 years of age, as previous studies have suggested, we would expect to find that many of the participants will not have comprehended the color terms by 2;6. Study 2 extends this exploration to 11 other languages but for production only. Study 2 allows us to test whether color words are learned following a universal order from their biological roots. If so, the overall order in which color words are learned should show little variability across languages. Finally, Study 3 examines the impact of color word frequency and syllabic complexity of the different color word forms across these languages, in attempt to identify potential sources of variation. Similar to the goals of Study 2, if color words are learned purely based on their biological foundations, then frequency and complexity would not be expected to be strong predictors; alternatively if cultural and linguistic factors shape color word learning, then frequency and complexity should be important factors.

Study 1

Method

Participants

2,962 8- to 30-month-old participant's details were filled out by parents, either on paper or online before a visit to testing facilities either at the Plymouth Babylab or the Oxford BabyLab. Participant information that was lacking in either age or gender information was not included in the analysis. The majority of infants visited the laboratories only once, making these analyses cross-sectional rather than longitudinal in character. A small number of participants visited more than once, giving a total of 3,413 completed CDIs (1,653 female).

Materials

In Study 1, previously collected data from the Oxford CDI (Hamilton, Plunkett, & Schafer, 2000) were used to examine color word comprehension and production. The Oxford CDI is a British adaptation of the MacArthur-Bates CDIs (MB-CDIs; Fenson et al., 1994, 2007), measuring comprehension and production in 416 terms, and used from the earliest stages of word learning, up until around 30 months of age. The Oxford CDI contains four color terms: *red*, *blue*, *green*, and *yellow*.

Analysis

In this, and the following analyses, parental report data are modeled with Bayesian binomial models. The objective of this analysis is to fit a curve to data that is binomially distributed (yes/no data), and in doing so be able to calculate a developmental trajectory (for general frequentist examples of modeling, see Mirman, 2014, for some Bayes-specific examples, see Bürkner, 2017b). In addition to being able to view these trajectories, the population-level coefficients of the model provide information as to the effects that shape the model. In the approach used here, it can be considered strong evidence for a coefficient being an important factor in the model if the 95% credible interval of the coefficient does not intersect with 0 (Kruschke, 2013). This kind of model could be fitted with either a frequentist generalized linear model, or a Bayesian model. The choice of Bayesian analysis was made in order to make inference-based analysis on the model output while simultaneously avoiding shortcomings associated with some frequentist models (Cumming, 2014). In addition, the Bayesian method allows for greater flexibility in modeling, and the ability to fit complex models that maximum likelihood methods can fail to capture (Bürkner, 2017b).

The four color terms were isolated for each participant, and modeled with two separate Bayesian binomial models, each with four chains of 12,000 iterations, of which 2,000 were a warm-up. The chains were thinned by 2, to allow minimal autocorrelation. Both models included age, gender and

the color word in question as population-level (fixed) effects, as well as an interaction between age and gender to allow the possibility of different slopes for each gender. Color and gender, both categorical variables, were treatment coded, comparing to *blue* in the case of color, and to *female* in the case of gender. Age was treated for these analyses as a continuous, numeric variable. The data were modeled using brms (Bürkner, 2017a), running in rstan (Stan Development Team, 2016), using the code (for the comprehension data):

$$brmðKnownjtrialsðTrialsÞ~ Age þ Color$$
$$þ Gender þ Age : Gender, family ð1Þ$$
$$¼ binomialð^{0}logit^{0}ÞÞ$$

An identical model was run on the production data. Priors were largely uninformative Student's *t*-distributions with 10 *df*, a mean of 0, and deviance of 1 for all the population-level effects. Both models were checked for proper convergence with no divergent steps, and a $\hat{R}$ of < 1.1, with an effective sample size of > 10% of the total sample size, and a Monte Carlo standard error of < 10% of the posterior standard deviation.

Results

The fit of the binomial curves can be seen in Figure 1. Overall there is a clear difference between comprehension and production for each of the
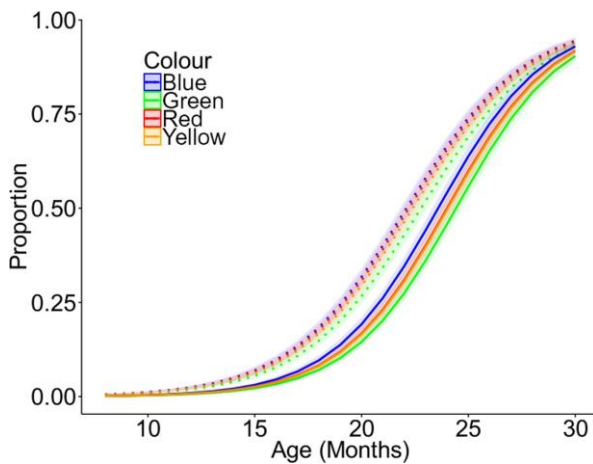

Figure 1. Results of two fitted Bayesian models to the Oxford Communicative Development Inventory comprehension and production data. Dotted lines indicate comprehension, solid lines indicate production. Narrow bands around each line indicate the credible interval of the mean.

colors, such that approximately 50% of infants are reported to comprehend, for example, *blue* at age 21 months, but only 25% of infants produce it at the same age. A similar difference between comprehension and production exists for each of the color words tested. Below the results of the comprehension data and production data are discussed separately, in detail.

The results of the fitted model to comprehension data in the Oxford CDI suggest that each of the color words are comprehended by around 50% of the infants at 21 months. *Blue* appears to be the first of the color words comprehended, by a small margin, whereas there is strong evidence that *green* is the last of the four color words to be comprehended. These findings are supported by the results of the model as seen in Table 1. The "Est." column gives the estimate of the means of the posterior distribution, whereas the two columns under "95% CI" give the upper and lower 95% Credible Interval around the estimate, and the "Err." column denotes the standard error on the estimate. "Samples" denotes the number of samples gathered for each individual parameter. Thus there is evidence for a difference between *blue* and *green*, where the 95% credible interval does not overlap with 0.

The results of the model also suggest a gap in comprehension between male and female participants. Notably, although there appears to be a possible early advantage for male infants comprehending color words, the interaction with age provides strong evidence of a shallower slope in learning for male infants than for female infants, indicating an overall advantage for girls.

The model of production of color terms in the Oxford CDI reveals a similar profile of learning to produce the terms to that of the comprehension data (Table 2). Again, there is evidence that *green* is produced slightly later than the other colors, most notably *blue*. For production, the 95% credible

Table 1

Results of Model on Comprehension of Color Words in the Oxford Communicative Development Inventory

	Est.	Err.	95% CI	Samples
Intercept	—8.40	.19	[—8.78, —8.03]	14,753
Age	.39	.01	[0.37, 0.41]	14,760
Green	—.25	.07	[—0.38, —0.12]	16,302
Red	—.04	.07	[—0.17, 0.08]	15,952
Yellow	—.11	.06	[—0.24, 0.02]	16,173
Male	.61	.26	[0.11, 1.12]	14,082
Age:male	—.04	.01	[—0.07, —0.02]	14,145

Table 2

Results of Model on Production of Color Words in the Oxford Communicative Development Inventory

	Est.	Err.	95% CI	Samples
Intercept	—9.28	.23	[—9.72, —8.84]	14,587
Age	.40	.01	[0.38, 0.42]	15,227
Green	—.34	.07	[—0.48, —0.20]	17,519
Red	—.17	.07	[—0.31, —0.03]	15,388
Yellow	—.18	.07	[—0.32, —0.04]	16,410
Male	—.54	.32	[—1.16, 0.07]	14,385
Age:male	.01	.01	[—0.02, 0.03]	14,326

interval for the difference between *blue* and each of the remaining three color words does not intersect with 0. There is no strong evidence for a difference in either baseline or slope between the two genders in the case of production.

Finally, a subset of these participants' caregivers were asked to fill out supplementary information to the Oxford CDI. This information asked caregivers to confirm whether the child comprehended or comprehended and produced all 11 basic   color terms (*red*, *blue*, *green*, *yellow*, *black*, *white*, *pink*, *orange*, *purple*, *brown*, and *gray*). A total of 256 participants completed the supplementary information. Table 3 shows the information by age group.

Identical models for comprehension and production to the previous models were fit to the data, in order to examine trajectories of all the 11 basic color words. The fit of the binomial curves can be seen in Figure 2. Although the sample size for this data is limited, both models fit the trends for comprehension and production from the full Oxford CDI data, and show the close contiguity from the first four terms (*red*, *green*, *yellow*, and *blue*).

The supplementary data to the Oxford CDI show that parents report *gray* to be comprehended and produced last, and *brown* to be learned moderately late, along with *black* and *white*. *Purple*, *pink*, and *orange* are produced and comprehended between the first four colors, and the later four. Overall the timing for learning these color terms seems to agree with the overall Oxford CDI data. Model coefficients can be seen in Supporting Information.

Table 3

Ages of Participants Who Completed Supplementary Oxford Communicative Development Inventory Data

Age	12	16	18	19	24	26	28	30
N	52	59	4	72	39	5	5	20

Study 2

Method

Participants

Data from 22,642 participants were downloaded from the Wordbank database http://wordbank.stanford.edu/ (Frank, Braginsky, Yurovsky, & Marchman, 2017) on November 18, 2016. Data were downloaded for 11 languages based on two selection criteria: First, the data needed enough participants to make it a generalizable sample, for the purposes of this experiment that was 600 participants. Second, the CDI data for that language needed to contain each of the six color terms being examined (the four in Study 1, plus *black* and *white*). Participants older than 2;6 were excluded from the analysis. Final participant numbers for each of the language groups can be seen in Table 4.

Materials

In each data set downloaded, participants' guardians had filled out the MB-CDI in the first language being developed by the child. Foreign language adaptations of the CDIs are often not direct translations of each of the terms, but instead are adapted to account for the differences in language. Each different language CDI set contained production data from participants between 1;4 and 2;6, with the exception of *Russian*, *German*, and *Italian* which began at 1;6, and *Swedish*, which contained data from participants at 3-month intervals from 1;4 to 2;4.

In English, the terms used were *red*, *green*, *yellow*, *blue*, *black*, and *white*. All the languages used in this study had corresponding words for these terms. Where multiple terms exist for one of these colors, the most common one was used. In *Cantonese*, the standard modern equivalents for these terms were used, whereas *Mandarin* was already translated into English, it is expected that the terms used were the same as used in *Cantonese*. In *Russian*, the word *siniy* (синий) was used for *blue*, as *goluboy* (голубой) was not included in the CDI. The full list of words used in the analysis can be viewed in Table 5.

Analysis

The data were modeled in a similar fashion to that of Study 1, except that the number of iterations was increased to 20,000, of which 4,000 remained as a warm up, to allow for the larger data set. Age, gender, and color were again population-level
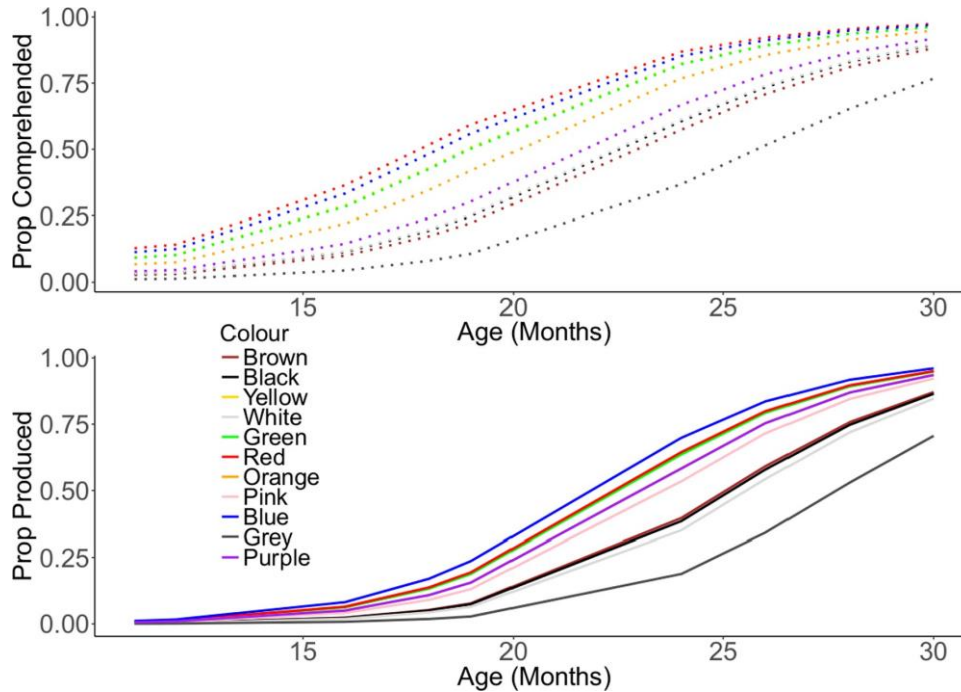
Figure 2. Comprehension and production data from a subset of Oxford Communicative Development Inventory participants, fitted with binomial curves. The trends for some color terms (e.g., production of *orange*), directly overlap with other terms. 95% CI are not displayed here.

effects, also including an interaction between age and color. Age and color were both nested within the group-level effect of language.

Priors for population-level effects were as in Study 1, priors on group-level standard deviations were default half *t*-distributions with 3 *df*, while priors on group-level correlations were default Cholesky factors. The brms model code was:

Table 4
Numbers of Participants in Each Language of the MacArthur Bates Communicative Development Inventory Surveys

Language	N
Cantonese	987
Danish	2,863
English (US)	5,450
French (Quebec)	827
German	1,183
Italian	639
Mandarin	1,056
Norwegian	6,931
Russian	712
Spanish	1,094
Swedish	900

brmðProducesjtrialsðTrialsÞ~ Age þ Genderþ Color þ Age:Color þ ðAge þ ColorjLanguageÞ; ð2Þ family ¼ binomialðlogitÞÞ

Results

Figure 3 shows the different trends in producing color words, dependent on the language being learned. The model again converged with $R^\char`^ = 1$ and no divergent transitions. The model coefficients (Table 6) show very strong evidence for an effect of age, as well as strong evidence for an effect of gender, with male participants being generally behind female participants.

With the exception of *white*, the model suggests there is convincing evidence for the other colors to be produced ahead of *black*, with the 95% CI on the difference between them and *black* not including 0.

This is consistent with the graphs depicted in Figure 3, where the four primary colors are produced before *black* and *white* in many of the languages examined. Overall, the general trend is a close contiguity with the four chromatic color words, and then again a close contiguity between the two achromatic words. In many cases *red* or *blue* are the

Table 5
Words Used in Each Language Based on MacArthur Bates Communicative Development Inventories (CDIs)

Language	Terms					
English	red	green	yellow	blue	black	white
Cantonese	紅	綠	黃	藍	黑	白
Danish	rød	grøn	gul	blå	sort	hvid
French	rouge	vert	jaune	bleu	noir	blanc
German	rot	grün	gelb	blau	schwarz	weiß
Italian	rosso	verde	giallo	blu	nero	bianco
Mandarin[a]	red	green	yellow	blue	black	white
Norwegian	rød	grønn	gul	blå	svart	hvit
Russian	красный	зеленый	желтый	синий	черный	белый
Spanish	rojo	verde	amarillo	azul	negro	blanco
Swedish	röd	grön	gul	blå	svart	vit

[a]Mandarin CDI data were made available in English.

first terms learned, consistently above *yellow* and *green*.

In the five Germanic languages examined (English, German, Danish, Norwegian, and Swedish), there is a close contiguity in the production of the four primary colors, with green possibly the last of those four colors, except in the case of German. In these languages, each of the color terms has been produced by at least 75% of infants at 30 months.

Figure 3 shows that the time-course of color word production in Romance (French, Italian, and Spanish) languages is not as uniform as it is in the Germanic languages. In French, color word learning happens relatively early, with all six color terms known by around 75% of infants by 30 months. In Italian, color word learning is slightly later than it is in French, whereas in Spanish, color word learning happens much later, with each color word only produced by around 50% of participants tested at 30 months. The most notable feature of the Romance languages examined here is that although black and white tend to follow behind the four primary colors, as with the Germanic languages, they do not do so by as large a margin as in the Germanic languages. Black is consistently the last word produced in these three languages, albeit by a small margin. The other consistent aspect in the Romance languages is that green tends to be produced after the other three primary colors, closer in timing to white.

Within the two Sinitic languages (*Mandarin* and *Cantonese*), there are distinct differences in the timing and order of color words. Although a dominant feature of the two Chinese languages is very early production of the word for *red*, in Mandarin this is matched by *white* as one of the first color words

produced, which is not the case in Cantonese. In Cantonese, the other five color words are produced at essentially the same rate, whereas in Mandarin, *blue* is produced around a month later than the remaining three color terms. There is a large difference in the rate of color word learning between these two languages as well. In Mandarin, the parents report that almost all the six color terms are produced and understood by almost all infants by 30 months of age. In contrast, in Cantonese, most color words are produced by about 60% of infants by 30 months, with the exception of red, which is known by around 75% of infants.

In Russian, the overall pattern of color word learning is not dissimilar to that of the Romance languages. The majority of color words, excluding *white*, are produced by around 70% of infants by 30 months, and *black* is produced after the primary colors, although not by the amount seen in the Germanic languages. Where Russian differs greatly is that *white* is produced long after the other color terms, with a gap of around 3 months. By 30 months of age, only around 30% of Russian infants are reported to produce the word *white*. In *Russian*, the last of the primary colors to be produced is *yellow*, reflecting another possible difference in ordering.

The model presented here shows substantial differences in the order of color words, and in the timing with which color words are produced. Although this model indicates that there is not a universal order in color word learning, suggesting that there are cultural and linguistic differences, it does not yet answer what factors cause this difference—the motivation for the third study.
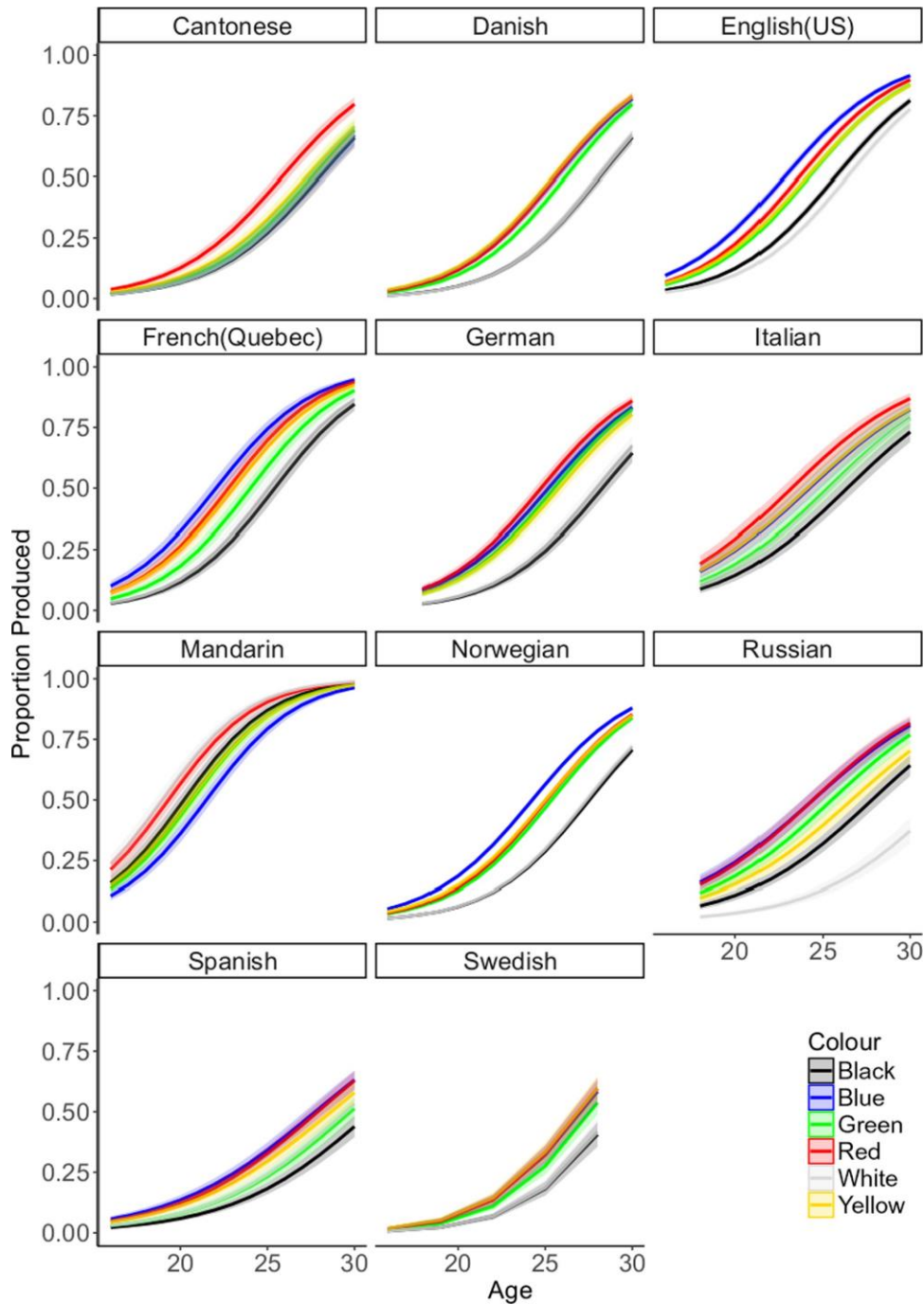
Figure 3. MacArthur-Bates Communicative Development Inventory production data modeled with a binomial Bayesian model, separated by language. Bands around each line indicate confidence in the mean.

Study 3

Method

Materials

For Study 3, two sets of data are included. The first is the same set of MB-CDI data used in Study 2, whereas the second is data on the frequency of occurrence of words. Frequency data were obtained by downloading CHILDES (MacWhinney, 2000) CHAT transcripts from http://childes.talkbank.org/ on December 19, 2016 and examining the frequency of color word appearance in each of the languages tested. Frequencies were then tallied up for

each color term and divided by the total number of words to yield a proportion (see Figure 4). In line with Goodman, Dale, and Li (2008), frequency of input was approximated by counting the frequency of appearance of each color word in each language in CHILDES when spoken by the mother. Frequency varies substantially between languages and colors.

It has previously been shown that frequency, category size (i.e., how many shades are encompassed by the single term) and perceptual salience (i.e., the Euclidean distance from gray at the center of the space) can predict precise color word learning in a behavioral task (Yurovsky, Wagner, Barner, & Frank, 2015). In this study, only frequency is used

as a predictor, of those three possible options. Although it is extremely likely that category size and perceptual salience are useful predictors, data on this for all the languages included in this study are not currently available.

In each language, only transcripts in which the infants were 3 years or younger were used and only occurrences in which the meaning of the color term could be understood by native speakers to refer to the term, and not idiomatic expressions were accepted. For many of the European languages, care was taken to ensure that color terms in each separate gender were included. Compound words, where the noun is made up of a color word and another word (e.g., *blueberry*), were not included in the count, particularly in the case of the Sinitic languages, where the compound word may not require an understanding of the meaning of the color words involved. As an example the term *hóng lǜ ɗeng* (lit. red green light, meaning traffic light) was not included as one could understand the meaning of the term without necessarily understanding the words referring to *red* and *green*.

In two cases, because the color term was a homonym with another commonly used term, or transcribed the same way (German *white* and *know*), the frequency had to be predicted. In the German case, the frequency of *Weiß* was calculated by working out the ratio of masculine to feminine endings of each of the color terms, and multiplying the average of that by the amount of times the feminine *Weiße* was used.

Table 6

Main Population-Level Effects on Fitted Model of MacArthur Bates Communicative Development Inventory Data in 11 Languages

	Est.	Err.	95% CI	Sample
Intercept	—8.84	.47	[—9.76, —7.91]	21,363
Age	.34	.02	[0.3, 0.38]	21,851
Male	—.4	.01	[—0.43, —0.37]	24,205
Blue	.98	.24	[0.5, 1.45]	25,532
Green	.48	.21	[0.07, 0.89]	26,713
Red	.83	.2	[0.44, 1.22]	27,062
White	—.11	.26	[—0.62, 0.41]	27,915
Yellow	.88	.21	[0.47, 1.3]	26,174

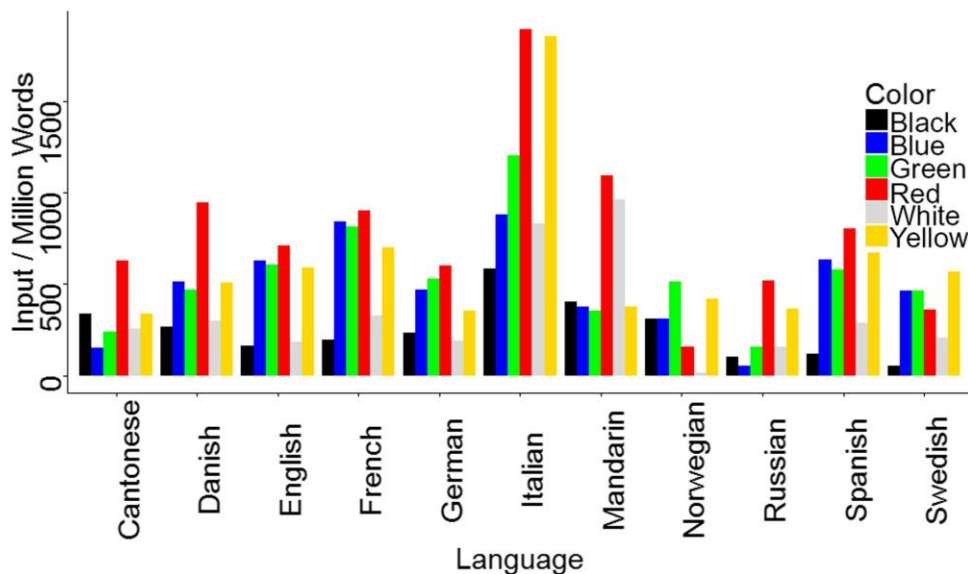*Note.* Colors are as compared to black. Interaction terms and group-level terms are not included.

Figure 4. Input frequency of occurrence of color words for infants up to 3 years of age, by language.

In both Mandarin and Cantonese, care was taken to ensure that appearances transcribed in both pinyin and characters were included.

Analysis

In this study, Bayesian models were constructed in a similar fashion to the previous studies, with the same priors. In contrast to Study 2, the categorical variable of color was replaced with the numeric variable of frequency, which was multiplied by one thousand to appear on the same scale as the other variables. An additional variable of syllabic complexity was also added, which was calculated as the number of syllables of each color term. Because frequency is nested within language, each language will be affected differently by the coefficient of frequency, which allows for greater flexibility in fitting the model, but also means the model fit is less affected by the discrepancy in the overall frequency numbers in each language. The model used the log of the frequency, due to evidence that frequency of input should be log transformed (Anderson & Schooler, 1991; Yurovsky et al., 2015).

Three models were run, each with the same specifications as in Study 2, except for increased iterations to 24,000 and warm-up of 8,000 iterations, in order to allow for the more complex structure of the model. Running three separate models allowed the assessment of the addition of each term into the first, base model. The model codes used were as below:

$$\text{brm}(\text{Produces}|\text{trials}(\text{Trials}) \sim \text{Age} + \text{Gender} + (\text{Age}|\text{Language}); \text{family} = \text{binomial}(\text{logit})) \tag{3}$$

$$\text{brm}(\text{Produces}|\text{trials}(\text{Trials}) \sim \text{Age} + \log(\text{Frequency}) + \text{Gender} + (\text{Age} + \log(\text{Frequency})|\text{Language}); \text{family} = \text{binomial}(\text{logit})) \tag{4}$$

$$\text{brm}(\text{Produces}|\text{trials}(\text{Trials}) \sim \text{Age} + \log(\text{Frequency}) + \text{Complexity} + \text{Gender} + (\text{Age} + \log(\text{Frequency})|\text{Language}); \text{family} = \text{binomial}(\text{logit})) \tag{5}$$

Results

The three models in this study were analyzed separately, and then compared with leave-one-out

information criteria (LOOIC). The first model analyzed color word production using only age and gender as population-level effects, with a group-level effect of age varying by language (LOOIC = 12,808.44, *SE* = 264.22). In this first, base model, there was strong evidence for an effect of age (95% CI [0.29, 0.36]) and for an advantage to female participants (95% CI [-0.41, 0.36]).

The second model added frequency of the appearance of each color word in each language in the CHILDES database as a predictor. Using frequency, age and gender to model the MB-CDI color word data resulted in a dramatic improvement over the basic model containing only predictors of age and gender (LOOIC difference = 1,871.41, LOOIC *SE* of difference = 157.37). There was strong evidence that frequency (95% CI [0.30, 0.76]), age (95% CI [0.29, 0.37]), and the gender difference (95% CI [-0.42, 0.37]) all predicted word learning, with the 95% credible interval not intersecting 0 for any of those predictors.

In the final model, color words in each language were assessed on the number of syllables each possessed, and added as a population-level effect into the previous model. Syllabic complexity was again found to be a potential but weak factor in predicting color word learning (95% CI [-0.11, 0.00]), whereas frequency, age, and gender continued to have strong predictive power (Table 7). This final model proved to be arguably a slightly better fit than the frequency-only model (LOOIC difference = 1.06, *SE* difference = 5.17). The final model was successful at capturing much of the variance of color word learning in different languages but was not as optimal as the original descriptive model of the data which used color terms as categorical variables in Study 2 (LOOIC difference = 1,682.46, *SE* difference = 177.24). In addition, the presence of positive evidence for slope of frequency to differ in each language (95% CI [0.22, 0.64]) suggests that the effect that frequency has differs greatly in each language.

Table 7

Population-Level Effects of Final Predictive Model Using Both Frequency and Syllabic Complexity as Predictors of Color Word Learning

	Est.	Err.	95% CI	Sample
Intercept	—7.68	.39	[—8.45,—6.91]	4,747
Age	.33	.02	[0.29,0.37]	4,251
Male	—.39	.01	[—0.42,—0.37]	22,618
Log(Frequency)	.54	.12	[0.3, 0.77]	6,026
Complexity	—.06	.03	[—0.11, 0]	22,695

One of the main points of difference between the predicted model and the original descriptive model was that the learning of *blue* was constantly underestimated in the final predicted model. Languages such as English, where the data suggest a clear advantage to *blue,* are instead modeled to show *blue* coming in behind other languages. In addition the model predicts a closer contiguity between the color terms than is realized in the data. These factors suggest that as well as the obvious strong effect of frequency, there are other elements as well, such as the category size or perceptual salience of the color (Yurovsky et al., 2015).

Discussion

The present research uses measures of color word learning from parental reports to assess the time line and trajectory of color word learning from around 15- to 30-months of age. Recent research has substantiated the existence of a biological component in color category formation in infancy (Skelton et al., 2017) but raises the question: What causes the shift to adult color categories? This strong biological contribution to early color categories also reignites discussion of a potential universal order in which color words are learned (Pitchford & Mullen, 2002), analogous to that in which color words are proposed to emerge historically in languages (Kay et al., 2011).

The present research demonstrates that the order in which color words are produced varies greatly between languages. The results from Study 2 provide strong evidence for differences in color word learning around the world, albeit with many similarities between groups that share a similar language and culture. Study 1 demonstrates that the onset of comprehension of color terms follows a very similar trajectory to that of production, even for color terms learned later, such as *brown* and *gray*. Whereas Study 2 only measures production, comprehension can be assumed to precede it by a similar margin as shown in Study 1. Finally, in Study 3, the results of Study 2 are successfully approximated by modeling using the frequency of input and the lexical complexity of the color term. Based on the evidence of Study 3, it appears that the frequency with which infants hear a word, and the syllabic complexity of the word, are strong predictors in the timing of color word production. This suggests that the timing of color word learning is very much a linguistic and culturally mediated process.

That color word learning is not universal, despite the biological foundations of color categories, suggests a change in process in the understanding of color categories. The visual color categories evidenced in prelinguistic infants (Franklin, Clifford, Williamson, & Davies, 2005; Skelton et al., 2017) must adapt with the slow comprehension of color words and their meaning (Franklin, Drivonikou, Clifford, et al., 2008). This process is not universal; the scope of the category for each word varies by language (Roberson et al., 2008), as does the timing of learning the word. Infants, in the learning of a color term, are taught by frequent exposure to the term, as seen in Study 3. Thus, their understanding shifts as they slowly grasp a full comprehension of the meaning of the term. This may come about earlier or later, depending on how often they are exposed to the term. Wagner et al. (2013) found that when infants first comprehend a color word, they comprehend the category center but overextend it to include other colors. As the infant comprehends more terms, the additional category centroids force the infant to update their understanding of the original color category, shrinking the category boundary with the addition of more terms. Thus a partial comprehension of the color word precedes production, but is slow to mature. The comprehension data presented here, showing that color words are learned differently due to different cultural and linguistic settings, capture the earliest part of that process—the basic comprehension of the focal color term.

This study points to color word learning in general occurring much earlier than previously reported (Heider, 1971; Mervis et al., 1995; Pitchford & Mullen, 2002; Shatz et al., 1996; Soja, 1994). Although it is possible that this may be part of a general trend of children learning color words earlier than they used to (Franklin, 2006), there are two other considerations. A major consideration is that by asking parents to record whether their children can produce these terms, a larger-scale picture of word production that may be more sensitive than laboratory studies has been obtained, partly through the size of the samples used. The other consideration is that we may be measuring an earlier process, as it is possible that parents are able to report an early comprehension or production that is not yet consistent, or that the child cannot yet confidently reproduce in front of a stranger in a laboratory. The model of the Oxford CDI data (Study 1) demonstrates that toddlers' efforts to *understand* the meaning of color words takes place earlier again, possibly during the second year of life in many languages. This again suggests that color word

learning may not be as difficult as previously thought (Andrick & Tager-Flusberg, 1986; Kowalski & Zimiles, 2006; Soja, 1994). Although this may, in part, be due to environmental factors that promote the usage of color words with young children, such as a focus on color terms by the parents or an increase in colored plastic toys around the home, it is more likely reflective of a difference in measurement sensitivity. Parents are very sensitive to infants understanding and production of specific words (Hidaka & Smith, 2010) and have the opportunity to see them comprehend and produce words in a variety of contexts. In contrast, an early comprehension of color words is much harder to assess, unless explicitly tested, and the contexts in which they are used are far more limited, so the earliest point of comprehension of color terms is much more difficult to ascertain (Ramscar, Thorpe, & Denny, 2007). It is likely that the patterns of color word learning are analogous to those of other classes of words, where infants have a basic comprehension, which is then refined slowly over time after they start producing the word. The toddler's comprehension of the boundary separating color terms is slowly refined as they are exposed to more colors and color terms, in the same way as they learn to distinguish between two similar categories (Wagner et al., 2013). It may simply be that frequency affects learning of color words in the same manner as other classes of words.

In Study 1, the gap between comprehension and production is shown to be quite small and can be assumed to be similar for Study 2. This does, however, raise the question of measuring comprehension using parental report. Parental reports of production data are likely to be largely accurate. Asking parents to assess comprehension of a word, however, has been criticized (Houston-Price et al., 2007; Tomasello & Mervis, 1994) as an accurate method, although other findings have shown that CDI comprehension measures are useful estimates, at least in the case of concrete nouns (Bates et al., 1988; Mills et al., 1993, 1997). The short gap between comprehension and production was shown again for all 11 colors in the subset of participants who completed the additional color word survey. Further suggestions that, if anything, comprehension estimates in CDI studies are an underestimation (Styles & Plunkett, 2009), suggest that the short gap between comprehension and production indicates the possibility of earlier comprehension than reported here.

Although there is great variety in the timing and order of color word acquisition, as demonstrated in Study 2, there are also overall trends. In general,

the data from Study 2 indicate that *blue, red,* and *yellow* are learned before *green*, which is learned before *black* and *white*. This may reflect the basic biological underpinnings of color categories (e.g., Skelton et al., 2017), particularly if the order of color category learning is underpinned by factors such as perceptual salience (Yurovsky et al., 2015). Additionally, it is possible that these overall trends reflect the fact that the majority of languages chosen for this study are European languages, and a wider selection of languages may show a slightly different pattern. In Mandarin, for example, *white* is one of the first color words produced, unlike any of the European languages in this study.

In Study 3 it is demonstrated that much of the variance in the difference in timing of the acquisition of color words occurs due to input frequency and syllabic complexity. The variance in input frequency (Figure 3) is incredibly large, a factor that may be cultural, or a peculiarity of the data. The prevalence of *red* in the Sinitic languages, for example, is likely to be cultural, given the associations between that color and luck, and fortune. Although this is likely true in many cases, the recordings of child–parent interactions that make up CHILDES are limited and could be biased by, in some instances, an infant playing with a toy of a particular color, or a color-based game. The variation that can be seen in the frequency of input data from CHILDES attests to this. Italian showed a much higher frequency for most colors than the other languages, suggesting a possible activity bias in one or more of the data sets. Frequency of input is a powerful predictor for color word learning but does not account for all the variance. Syllabic complexity appears to account for some further variance, suggesting that the length of a word may make it harder for infants to learn. One possibility is that other predictors, such as the visual salience of the color and the size of the color word category could account for some of the remaining variance (Yurovsky et al., 2015). In this sense, it is likely that the absence or present of additional color terms (such as the additional term for *blue* in Russian) may play an important role in the timing of color word learning, as they change the category size for each of the surrounding colors.

This study addresses the start of the color word learning process in infants. However, this study is limited by the materials that are available to examine this phenomenon. The 11 languages selected for use in this study do not necessarily give a complete picture of color word learning around the world, as materials are not available for some widely spoken languages, such as Arabic. Additionally, the lack of

data available on the perceptual salience and category size of each color word in all the languages prevent us from obtaining the complete picture suggested by Yurovsky et al. (2015). Finally, Study 3 relies on rough measurements of frequency, adding considerable noise to the data. That such a measurement of frequency of input still strongly predicts timing of acquisition demonstrates the strength of this effect.

It should be stressed that what is being measured here is not necessarily an adult-like understanding of color words by young toddlers but rather the beginning of a slow process of establishing the contents of a color word category. Although they may still be prone to errors in applying those terms correctly (Pitchford & Mullen, 2003), they may have understood that the color term refers at least to the focal area of that color word category. Infants clearly begin to understand color words much earlier than first thought, and they do so with great variety, depending on both the individual and the language which they speak.

Conclusion

This study provides strong evidence for cultural and linguistic variation in the formation of color categories, through analysis of parental surveys of British English children, and matched by parental surveys from around the world. Color word learning follows no universal pattern or timeline but instead varies with differing languages and cultures. In this sense, and in the sense that a partial comprehension seems to precede production (Wagner et al., 2013), color words seem to be learned in much the same way as any other class of word. The results also suggest that color word learning may occur much earlier than previously seen, thus suggesting that perhaps color word learning is not as uniquely difficult as had previously been assumed (Franklin, 2006; Soja, 1994). Color word learning in this study was measured with parental report, and further behavioral investigations into color word production and comprehension will be crucial to fully understanding this topic.

Infant color categories appear to possess a biological, universal foundation (Skelton et al., 2017). However, the infant color categories change into adult-like understandings of the color terms, a process that, from the evidence presented here, is determined by the nature of the language which they learn. Despite the universal, biological origin of color categories, there is still an undeniable place for the cultural and linguistic.

References

Abbott, J. T., Griffiths, T. L., & Regier, T. (2016). Focal colors across languages are representative members of color categories. *Proceedings of the National Academy of Sciences of the United States of America*, *113*, 11178–11183. https://doi.org/10.1073/pnas.1513298113

Anderson, J. R., & Schooler, L. J. (1991). Reflections of the environment in memory. *Psychological Science*, *2*, 396–408. https://doi.org/10.1111/j.1467-9280.1991.tb00174.x

Andrick, G. R., & Tager-Flusberg, H. (1986). The acquisition of colour terms. *Journal of Child Language*, *13*, 119–134. http://doi.org/10.1017/S0305000900000337

Bates, E., Bretherton, I., & Snyder, L. (1988). *From first words to grammar: Individual differences and dissociable mechanisms*. New York, NY: Cambridge University Press.

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

Bornstein, M. H. (1985). On the development of color naming in young children: Data and theory. *Brain and Language*, *26*(1), 72–93. https://doi.org/10.1016/0093-934X(85)90029-X

Brown, R., & Lenneberg, E. (1954). A study in language and cognition. *Journal of Abnormal and Social Psychology*, *49*, 2454–2462. http://dx.doi.org/10.1037/h0057814

Bürkner, P.-C. (2017a). brms : An R package for Bayesian generalized linear mixed models using Stan. *Journal of Statistical Software*, *80*. https://doi.org/10.18637/jss.v080.i01

Bürkner, P.-C. (2017b). *Bayesian distributional non-linear multilevel modeling with the R package brms*. ArXiv ID: 1705.11123.

Cohen, L. B., Chaput, H. H., & Cashon, C. H. (2002). A constructivist model of infant cognition. *Cognitive Development*, *17*, 1323–1343. https://doi.org/10.1016/S0885-2014(02)00124-7

Cumming, G. (2014). The new statistics: Why and how. *Psychological Science*, *25*(1), 7–29. https://doi.org/10.1177/0956797613504966

Dale, P. S. (1991). The validity of a parent report measure of vocabulary and Syntax at 24 months. *Journal of Speech, Language and Hearing Research*, *34*, 565–571. https://doi.org/10.1044/jshr.3403.565

Dale, P. S., Bates, E., Reznick, J. S., & Morisset, C. (1989). The validity of a parent report instrument of child language at twenty months. *Journal of Child Language*, *16*, 239–249. https://doi.org/10.1017/S0305000900010394

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., . . . & Stiles, J. (1994). Variability in Early Communicative Development. *Monographs of the Society for Research in Child Development*, *59*(5), 1–185. https://doi.org/10.1111/j.1540-5834.1994.tb00169.x

Fenson, L., Marchman, V. A., Thal, D. J., Dale, P. S., Reznick, J. S., & Bates, E. (2007). *MacArthur-Bates Communicative Development Inventories: User's guide and technical manual* (2nd ed.). Baltimore, MD: Brookes.

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2017). Wordbank: An open repository for developmental vocabulary data. *Journal of Child Language*,

44(3), 677–694. https://doi.org/10.1017/S030500091600 0209

Franklin, A. (2006). Constraints on children's color term acquisition. *Journal of Experimental Child Psychology*, *94*, 322–327. https://doi.org/10.1016/j.jecp.2006.02.003

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

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

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

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

Goodman, J. C., Dale, P. S., & Li, P. (2008). Does frequency count? Parental input and the acquisition of vocabulary. *Journal of Child Language*, *35*(3), 515–531. https://doi.org/10.1017/S0305000907008641

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

Heider, E. (1971). "Focal" color areas and the development of color names. *Developmental Psychology*, *4*, 447. http://doi.org/10.1037/h0030955

Heider, E. (1972). Universals in color naming and memory. *Journal of Experimental Psychology*, *93*, 10–20. https://doi.org/10.1037/h0032606

Hidaka, S., & Smith, L. B. (2010). A single word in a population of words. *Language Learning and Development*, *6*, 206–222. https://doi.org/10.1080/15475441.2010.484380

Houston-Price, C., Mather, E., & Sakkalou, E. (2007). Discrepancy between parental reports of infants' receptive vocabulary and infants' behaviour in a preferential looking task. *Journal of Child Language*, *34*, 701–724.

Johnson, E. K., & Huettig, F. (2011). Eye movements during language-mediated visual search reveal a strong link between overt visual attention and lexical processing in 36-month-olds. *Psychological Research*, *75*(1), 35–42. https://doi.org/10.1007/s00426-010-0285-4

Kay, P. (2003). Resolving the question of color naming universals. *Proceedings of the National Academy of Sciences of the United States of America*, *100*, 9085–9089. https://doi.org/10.1073/pnas.1532837100

Kay, P., Berlin, B., Maffi, L., Merrifield, W., & Cook, R. (2011). *The world colour survey*. Chicago, IL: University of Chicago Press.

Kowalski, K., & Zimiles, H. (2006). The relation between children's conceptual functioning with color and color term acquisition. *Journal of Experimental Child Psychology*, *94*, 301–321. https://doi.org/10.1016/j.jecp.2005.12.001

Kruschke, J. K. (2013). Bayesian estimation supersedes the t test. *Journal of Experimental Psychology: General*, *142*, 573–603. https://doi.org/10.1037/a0029146

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk* (3rd ed.). Malwah, NJ: Erlbaum.

Mervis, C. B., Bertrand, J., & Pani, J. R. (1995). Transaction of cognitive-linguistic abilities and adult input: A case study of the acquisition of colour terms and colour-based subordinate object categories. *British Journal of Developmental Psychology*, *13*, 285–302. https://doi.org/10.1111/j.2044-835X.1995.tb00680.x

Mills, D. L., Coffey-Corina, S. A., & Neville, H. J. (1993). Language acquisition and cerebral specialization in 20-month-old infants. *Journal of Cognitive Neuroscience*, *5*, 317–334. https://doi.org/10.1162/jocn.1993.5.3.317

Mills, D. L., Coffey-Corina, S. A., & Neville, H. J. (1997). Language comprehension and cerebral specialization from 13 to 20 months. *Developmental Neuropsychology*, *13*, 397–445. https://doi.org/10.1080/87565649709540685

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

O'Hanlon, C. G., & Roberson, D. (2006). Learning in context: Linguistic and attentional constraints on children's color term learning. *Journal of Experimental Child Psychology*, *94*, 275–300. https://doi.org/10.1016/j.jecp.2005.11.007

Pitchford, N., & Mullen, K. T. (2002). Is the acquisition of basic-colour terms in young children constrained? *Perception*, *31*, 1349–1370. https://doi.org/10.1068/p3405

Pitchford, N., & Mullen, K. (2003). The development of conceptual colour categories in pre-school children: Influence of perceptual categorization. *Visual Cognition*, *10*(1), 51–77. https://doi.org/10.1080/713756669

Ramscar, M., Thorpe, K., & Denny, K. (2007). Surprise in the learning of color words. *Proceedings of the 29th Annual Cognitive Science Society*, *29*, 575–580.

Regier, T., Kay, P., & Cook, R. S. (2005). Focal colors are universal after all. *Proceedings of the National Academy of Sciences of the United States of America*, *102*, 8386–8391. https://doi.org/10.1073/pnas.0503281102

Regier, T., Kay, P., & Khetarpal, N. (2007). Color naming reflects optimal partitions of color space. *Proceedings of the National Academy of Sciences of the United States of America*, *104*, 1436–1441. https://doi.org/10.1073/pnas.0610341104

Roberson, D., Davidoff, J., Davies, I. R. L., & Shapiro, L. R. (2005). Color categories: Evidence for the cultural

relativity hypothesis. *Cognitive Psychology*, *50*, 378–411. https://doi.org/10.1016/j.cogpsych.2004.10.001

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

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

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

Sandhofer, C. M., & Smith, L. B. (1999). Learning color words involves learning a system of mappings. *Developmental Psychology*, *35*, 668–679. https://doi.org/10.1037/0012-1649.35.3.668

Shatz, M., Behrend, D., Gelman, S. A., & Ebeling, K. S. (1996). Colour term knowledge in two-year-olds: Evidence for early competence. *Journal of Child Language*, *23*, 177–199. http://doi.org/10.1017/S030500090001014X

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

Soja, N. N. (1994). Young children's concept of color and its relation to the acquisition of color words. *Child Development*, *65*, 918–937. https://doi.org/10.2307/1131428

Stan Development Team. (2016). *RStan: The R interface to Stan.* http://mc-stan.org/

Styles, S., & Plunkett, K. (2009). What is "word understanding" for the parent of a one-year-old? Matching the difficulty of a lexical comprehension task to parental CDI report. *Journal of Child Language*, *36*, 895–908. https://doi.org/10.1017/S0305000908009264

Tomasello, M., & Mervis, C. B. (1994). The instrument is great but measuring comprehension is still a problem. *Monographs of the Society for Research in Child Development*, *59*(5), 174–179. https://doi.org/10.1111/j.1540-5834.1994.tb00186.x

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

Yurovsky, D., Wagner, K., Barner, D., & Frank, M. C. (2015). Signatures of Domain-General Categorization Mechanisms in Color Word Learning. Proceedings of the 37th Annual Conference of the Cognitive Science Society, 2775–2780.

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's website:

Figure S1. Recruitment Information for Participants in the Oxford Communicative Development Inventory, by Age and Gender

Figure S2. Participant Information for All Data Used From the Wordbank Repository, Separated by Age and Gender

Figure S3. Fitted Final Model From Study 3, for Comparison to Study 2

Table S1. Comprehension Model From the Supplementary Data to the Oxford Communicative Development Inventory

Table S2. Production Model From the Supplementary Data to the Oxford Communicative Development Inventory

Table S3. Corpora Used in Analyzing Frequency of Occurrence for Each Color Word

Table S4. Full Frequency Calculations for Each Childes Corpus Used in Study 3