The Development of Color Terms in Shipibo-Konibo Children

Martin Fortier*[1], Danielle Kellier*[2], María Fernández Flecha[3], & Michael C. Frank[4]

[1] PSL Research University
[2] University of Pennsylvania
[3] Pontificia Universidad Católica del Perú
[4] Stanford University

Color word learning is an important case study for the relation between language and perception. While English color word learning is well-documented, there is relatively limited evidence on the developmental trajectory for color words, especially in languages from non-industrialized populations. We study color words and their acquisition in the Shipibo-Konibo (SK), an indigenous group in the Peruvian Amazon. In Study 1, we measure the color vocabulary in SK, updating data from the World Color Survey. We then study receptive and productive knowledge of color words in children, testing in both SK (Study 2) and Spanish (Study 3). Children learning the SK system show a protracted developmental trajectory compared with modern studies of English. Further, when children lack precise color term knowledge, they appeared to follow different strategies for SK and Spanish, using Spanish vocabulary in SK and overgeneralizing in Spanish. For both children and adults, bilingual vocabulary is used adaptively to facilitate task performance, broadly supporting communicative views of color vocabulary.

Keywords: Shipibo-Konibo; Color; Word learning; Bilingualism
Word count: 9258

Introduction

Color is where language and perception meet. Words such as "blue" and "red" draw boundary lines across a perceptually continuous space of hues. In English, there are 11 high frequency color terms that together span hue space, but this color categorization is not universal. For instance, Russian speakers use two distinct words to describe the colors light blue ("goluboy") and dark blue ("siniy"); and some languages have as few as two words [e.g., the Jalé people only have terms for "light" and "dark"; Berlin and Kay (1969)]. Why do languages vary in their color systems? One emerging consensus is that languages categorize the color spectrum in different ways in part due to functional demands (Gibson et al., 2017): both smaller and larger color systems are relatively optimal for different communicative needs (Regier et al., 2007; Zaslavsky et al., 2018).

Correspondence concerning this article should be addressed to Danielle Kellier*, Postal address. E-mail: danielle.kellier@pennmedicine.upenn.edu

Learnability is hypothesized to be one contributor to this cross-linguistic diversity (Chater & Christiansen, 2010; Culbertson et al., 2012). Some color systems may be easier to learn for children than others, or children may show inductive biases that shape color vocabulary. But the actual acquisition of color terms – while relatively well-studied in English (e.g., Sandhofer & Smith, 1999; Wagner et al., 2013) – is extremely under-studied across other populations. Berlin & Kay's seminal World Color Survey [WCS; Kay et al. (2009)] presented adult speakers of over 100 languages with differently colored chips and asked them to produce a label, characterizing the space of color vocabulary in a range of written and unwritten languages. The WCS is an invaluable resource for the cross-linguistic study of color vocabulary, but no comparable resource exists for cross-cultural studies of how this vocabulary is learned across childhood.

In the current project, our goals were (1) to characterize color term knowledge in an indigenous population previously studied by the WCS, the Shipibo-Konibo (SK), and then (2) to build on this foundation to characterize the developmental trajectory of color language acquisition in a group of children raised learning Shipibo-

Konibo, outside of the WEIRD (Western Educated Industrialized Rich Democratic) populations that are overrepresented in behavioral science (Henrich et al., 2010; Nielsen et al., 2017). In the remainder of the introduction, we review color vocabulary development in children, and then we turn to what is currently known about color terms in Latin American varieties of Spanish, such as Mexican, Colombian, and Bolivian Spanish, and in some Amazonian languages, such as Candoshi, Pirahã, and Shipibo Konibo. These two literatures set the stage for our own study.

The Development of Color Vocabulary

To adults, colors are extremely salient attributes of the perceptual world; even when color is seemingly task-irrelevant, we mention it (e.g., Sedivy, 2003). It is quite surprising then that children sometimes struggle to master color vocabulary. Early observations by Darwin, Bateman, Nagel, and others attest to individual children's delays in the correct use of color terms well into middle childhood; several diarists report 5 – 8 year olds with limited mastery of basic level color terms (reviewed in H.Bornstein, 1985). These observations are surprising in light of the body of infant research that suggests that infants' color discrimination abilities are relatively well-developed by the end of the first year of life (for review see e.g., Dobson & Teller, 1978).

Indeed, the age at which color words are learned has been shifting over the past hundred years, at least for English-speaking children. H.Bornstein (1985) documents substantial decreases in the age at which many children master their colors, citing four years as an age at which most children are proficient. In fact, this age may have even decreased further in the last thirty years, judging from recent studies (Wagner et al., 2013, 2018). What makes color words hard to learn, and why are they getting easier?

One prominent account of what makes color word learning difficult is that children may not recognize that color words pick out the perceptual dimension of hue at all (Bartlett, 1977; Sandhofer & Smith, 1999), and that once they do they then rapidly map colors correctly onto the appropriate range of hues in color space. This account nicely explains the observation that there is often a period during which children will produce an inappropriate color word when asked "what color is this?" – they know that color words go together and answer a particular question, they just don't know which color is which. A further point of parsimony for this account is that infants' color boundaries are not all that different in their placement from those of adults; thus, presumably the mapping task they face – from words to hues – is not all that difficult, once they recognize the dimension that they are attempting to map (Bornstein et al., 1976; Franklin et al., 2005).

On the other hand, when children's mapping errors are examined in detail, they show more systematicity than would be predicted by this account. In particular, Wagner et al. (2013) show that children who have not yet fully mastered the color lexicon nevertheless use colors in ways that are more consistent with overextension than with ignorance of the dimensional mapping – for example, using "blue" to refer to *blue* and *green* hues (which are close together in color space). These overextensions are reminiscent of noun overextensions that have been documented in early word learning, for example calling a horse "dog" (Eve V. Clark, 1973). Further, the order of acquisition for color word meanings in Wagner et al. (2013) was well-predicted by the frequency and perceptual salience of color categories (Yurovsky et al., 2015), supporting the view that color categories are learned gradually from perceptual experiences rather than all at once. Finally, both behavioral and eye-tracking evidence suggests that children show earlier comprehension than production for color words, a phenomenon seen throughout early word learning (Sandhofer & Smith, 1999; Wagner et al., 2018). In eye-tracking tasks, comprehension also shows evidence of perceptual overextensions, such that children fixate perceptually close distractor colors more than far distractors (Wagner et al., 2018). In sum, although attention to the dimension of hue may be one difficult component of color word learning, systematic mapping of words to particular regions of perceptual space is likely another.

Why is color learning occurring earlier in development, at least for English-learning children (H.Bornstein, 1985)? There are at least two obvious, plausible reasons. The first is the increasing prevalence of manufactured toys for children that vary exclusively in color (e.g., sets of plastic blocks of different colors) (Gibson et al., 2017). Such objects provide perfect contrastive input for mapping: if one is called "blue" and the other is not, such input implicates pragmatically that "blue" is an informative term (E. V. Clark, 1987; M. C. Frank & Goodman, 2014). The second is a cultural landscape for parents and early educators that presupposes color words are an important part of early childhood education practices, and as such should be taught explicitly (perhaps using toys specifically made for this purpose).

In the current paper, we ask about the trajectory of color word learning in an environment where both of these factors are less prevalent: that is, manufactured toys are less frequent, and parents are (at least anecdotally) far less motivated to provide color labels to their children. Here we are inspired by the work of Piantadosi et al. (2014), who studied the learning of number word meanings in children in an Amazonian culture. They found that, despite differences in developmental timing, the patterns of generalization of number meaning were generally similar to those documented in WEIRD populations. We are interested in whether we observe similar dynamics in color word learning. In the next section, we turn to the question of adults' color vocabulary in Spanish and Amazonian language, setting the stage for our studies of acquisition.

Color in Latin American varieties of Spanish and Amazonian languages

In their seminal work, Berlin and Kay (1969) established a framework for cross-linguistic differences in color vocabulary. They focused their work on basic level color terms (BCTs), the words that are highest-frequency and most consistently used when speakers of a language refer to visual hue. According to these authors, there is a fixed evolutionary sequence of stages that languages go through as they increase their color vocabulary; in this sense, if a language encodes a category from a particular stage, it must also encode those corresponding to all previous stages. So, for example, a Stage II system would add the term "red" to the colors already present in Stage I ("black" and "white"). It wouldn't be possible for a system to have "red" if it doesn't already have "black" and "white."

Although the original Berlin and Kay (1969) framework has been revised and questioned in subsequent work (e.g., Levinson, 2000), it continues to shape the research landscape on color. Yet there has been significant controversy about the applicability of the framework to Amazonian languages, specifically centered around the status of ad hoc color terms. Ad hoc terms are descriptors of objects or properties that are adopted for the description of hue (e.g., the use of terms like "blood" or "bloody" to refer to red objects). Such ad hoc terms are a common way that languages supplement color vocabulary (e.g., Kristol, 1980), with historical case studies suggesting that they can often become conventionalized BCTs [e.g., the English color "orange" derives from an ad hoc term based on the fruit; St. Clair (2016)]. As we will review below, the Berlin and Kay framework appears to apply relatively well to Spanish dialects, although some ad hoc terms are sometimes attested. In contrast, in Amazonian languages, ad hoc terms are more common and may make up a large proportion of the color language being used.

Only a handful of studies have explored the use of color terms in the varieties of Spanish in Latin America. Berlin and Kay (1969) examine the case of the Mexican dialect of Spanish, which they consider to be in Stage VII of their classification (color systems in this stage, the most advanced one, consist of between 8 and 11 color terms). They identify the following BCTs in Mexican Spanish: "blanco" (white), "negro" (black), "rojo" (red), "verde" (green), "amarillo" (yellow), "azul" (blue), "café" (brown or coffee-colored), "morado" (purple), "rosa" (pink or rose), "anaranjado" (orange, strictly referring to the color) and "gris" (gray). Monroy and Custodio (1989) offered some further information on Colombian Spanish based on materials collected for the Linguistic-ethnographic Atlas of Colombia, presenting some examples of ad hoc color terms referring to colors through objects prototypically instantiating these colors (e.g., vegetables, animals, food, metals, precious stones, fire and its derivatives, and atmospheric phenomena).

More recent work on Spanish largely confirms the WCS classification, while adding some dialectal nuance. Aragón (2016) offers an ethnolinguistic study of color terms in Mexican Spanish: she analyzes the elaboration of these meanings in dictionaries, as well as the references and associations that informants resort to for their own definitions. Aragón concludes that the local natural and cultural referents constitute a point of consensus among Mexicans when defining terms of color, even though these colors still follow the general schema of BCTs. Lillo et al. (2018) generally confirm these observations, finding an additional BCT in Uruguayan Spanish, "celeste" (sky blue), which we also observe in our study. This observation is also confirmed by Gibson et al. (2017) for Bolivian Spanish, who document 11 modal color names including "celeste" but not "gris" (gray).

Turning now to Amazonian languages, SK color terms were studied in the original WCS. In this original data collection effort, they list 21 distinct terms (though this could categorized as 20 as "huiso" and "wiso" are alternative spellings). seem like alternate spellings). As their protocol has the field experimenters ask only for BCTs, it is assumed that all recorded terms are basic, but only six terms appear in >5% of WCS trials; 10 terms appear in <1% of trials (see Figure 1A for a representation of this data). Immediately the issue of ad hoc terms rears its head (Levinson, 2000).[1] It is worth men-

[1] In fact, a greater diversity of color terms beyond the basic level is used in the data for the majority of WCS languages (Gibson et al., 2017, Figure S1), suggesting that the effort to elicit only BCTs in WCS was not successful.

tioning that two anthropological studies (Morin, 1973; Tournon, 2002) have also investigated the color terms used in this Amazonian language. However, these two studies contain some serious methodological pitfalls: a very limited number of color chips were tested with only a few participants. As a result, we will not further discuss these studies in the remainder of this article and will only focus in our study on a comparison with the WCS data.

To our knowledge, relatively little work has looked at effects of bilingualism in this space. But based on their work with Tzeltal participants (both Tzeltal monolinguals as well as Tzeltal-Spanish bilinguals), Berlin and Kay (1969) report that bilingualism did not skew their results regarding the existence of semantic universals in the domain of color vocabulary. Tzeltal has five BCTs: "?ihk´" (black), "sak" (white), "cah" (red), "yaš" (green) and "k´an" (yellow). This language is estimated to be transitioning from Stage IV to V, which is reflected in the ambiguity of the focus of "yaš" (grue, a category covering English green and blue hues). While all Tzeltal speakers acknowledge that "yaš" includes two major perceptual centers ("green" and "blue"), they vary in terms of their favored focal (either in the "green" or "blue" area). The authors posit that a long history of contact with Spanish has probably accentuated this pattern, and suggest that exposure to Spanish in schools will eventually cause "yaš" to be entirely restricted to greens, and "azul" (or some other Spanish term) will be adopted into the Tzeltal color system.

Several other indigenous Amazonian color systems were studied in the WCS. One of them, Candoshi, has been further examined by Surrallés (2016). Contrary to the WCS, Surrallés argues that no proper color terms exist in this language. If the fieldworkers of the WCS found otherwise, he claims, it is only because they misidentified the elicited terms as BCTs when they are nothing more than a series of ad hoc terms referring to objects or animals of the surrounding environment. For example, in Candoshi, the word for yellow is "ptsiyaromashi" ("like the feathers of a milvago bird"), the word for red is "chobiapi" ("ripe fruit"), the word for green is "kamachpa" ("unripe fruit"), etc. These findings lead Surrallés to argue that the Candoshi do not have a proper color system. When they use "color terms" they are not trying to subsume objects of the world under abstract color categories, but they are rather establishing horizontal and ad hoc comparisons between similar objects of the world.

A similar criticism of the WCS approach was given by Everett (2005) based on his study of Pirahã, another Amazonian language. Everett also rejected the idea that there are BCTs, arguing that the four color terms identified as basic in the WCS are not such. For example, the word identified as the BCT for *red/yellow* in Pirahã ("bi i sai") was argued to be simply a property descriptor meaning "blood-like." The argument here is that Pirahã color terms might be ad hoc comparisons rather than proper basic terms, though there was no quantitative evaluation of this claim such as analysis of the variability in term use.

Finally, Gibson et al. (2017) compare their Bolivian Spanish data with Tsimane, a language of the Amazonian piedmont. Out of a total of 80 color chips, Tsimane exhibited 8 modal or basic color terms. However, in their free-choice paradigm, Tsimane speakers showed greater variability in nearly all the color terms used for all color chips presented in their study. Thus, Tsimane speakers appear to show substantial ad hoc term usage.

In sum, while some dialectical differences can be noticed across varieties of Spanish, these slight variations are consistent with the general framework proposed by the WCS. Also consistent with the WCS, Amazonian color systems are characterized by fewer BCTs than the Spanish dialect systems. There is less consistency in the finding that ad hoc terms appear to play a central role in Amazonian color systems – and possibly also in some South American Spanish dialects (such as Colombian Spanish).

The Current Study

The Shipibo-Konibo people are an indigenous group located within the Peruvian Amazon. They are mainly horticulturalists, fishermen, occasionally hunters but are noted for their strong display of tradition despite increasingly regular interactions with the western world. They are also skilled traditional artists or artisans, resorting to these activities as a way to earn an income for their household. Their children receive formal schooling for 4 hours a day, both in SK[2] and Spanish. The amount of Spanish input they receive at school increases towards adolescence when they enter secondary education. There can be variation in how both languages coexist in the school setting from one village to another. Most SK adults are considered SK-Spanish bilinguals to different degrees although the elders may have only a minor grasp on Spanish.

[2]The phonemic inventory of SK language has 4 vowels (/i/, / /, /a/ and /o/) and 15 consonants: 3 plosives (/p/, /t/ and /k/), 2 affricates (/ts/ and / /), 2 nasals (/m/ and /n/), 5 fricatives (/ /, /s/, / /, / /, / /), and 3 approximants (/w/, / / and /j/).

The SK indigenous people are particularly interesting for at least two reasons: they differ from samples usually studied by cross-cultural evolutionary psychologists (Apicella & Barrett, 2016). Indeed, evolutionary psychologists are particularly interested in the study of contemporary hunter-gatherers because they are believed to be a good model of our Pleistocene ancestors. By contrast, like most riverine Amazonian cultures, the SK culture is not based on hunting and gathering, but on horticulture, fishing, and to a limited extent, hunting.

Further, because of their location on the Ucayali River, one of the main tributaries of the Amazon, the SK culture has always been enmeshed in rich trading networks involving other indigenous groups of the Andes and the Lowlands (in pre-conquest times) as well as Mestizos and Westerners (in post-conquest times) (Lathrap, 1970). It would thus be mistaken to think of this culture as an "isolated" or "preserved" one. On the contrary, having been extensively exposed to numerous cultural influences, the SK culture has been constantly reworked and reshaped through the centuries. The first deep transformation in Shipibo-Konibo culture can be traced to the XVIII century, when Shipibos, Konibos and Shetebos were forced to live together by Franciscan evangelization (Myers, 1974). Later, the second half of the 20th century was characterized by intense contact with the Spanish-speaking Mestizo populations established along the Ucayali River. As a result, today's SK culture straddles two worlds.

In Study 1, we examine the color vocabulary of current SK adults, comparing their vocabulary to results from the World Color Survey (a gap of more than 50 years). Next, we examine SK children's color vocabulary, focusing on their knowledge and their generalization of color terms across both SK (Study 2) and Spanish (Study 3). Through these three studies, we attempted to answer four primary research questions:

1. What is the color vocabulary of SK and how has it changed since the WCS data collection effort?
2. What is the developmental timeline of color term acquisition in a non-WEIRD population that has fewer industrial products (toys) and less early childhood education?
3. Is the developmental course – especially with respect to generalization and the dynamics of comprehension and production – similar to that which has been documented in studies of English color learning?
4. How is color term learning development affected by bilingual exposure?

To presage our conclusions, we find that SK color vocabulary has remained relatively consistent, with the exception of some intrusions from Spanish in areas of low coverage by the SK color system. Children learning the SK system show a protracted developmental trajectory compared with modern descriptive studies in WEIRD contexts. Further, when children lack precise color term knowledge, they appeared to follow different strategies for SK and Spanish: for SK, children fell back on Spanish knowledge, while for Spanish, we observed substantial over-generalization of terms (Wagner et al., 2013, 2018). Finally, we find that children draw on their Spanish knowledge especially for colors where there is high uncertainty among adult speakers, suggesting that they are adaptively using their bilingual knowledge to facilitate accurate naming.

Study 1

Before we could assess the developmental trajectory of color term knowledge in SK children, our goal was to replicate and update the characterization of the adult SK color system given by the World Color Survey. As the WCS study took place generations prior, we could not assume the SK color term mappings had remained static especially through years of industrialization and exposure to the Spanish language and its own color term system. As such, Study 1 used a modified version of the original WCS protocol, with an identical color chip set (subsampled to decrease task length). The goals were to characterize the current SK vocabulary and to generate a standard of adult knowledge against which subsequent child participants could be scored.

Methods

Participants. Our protocol for Study 1 and all subsequent studies received ethical approval from Pontificia Universidad Católica del Perú's Institutional Review Board. We recruited 39 adult participants (7 men). We experienced difficulty recruiting male participants as many of the men were away from the village during the day, resulting in a sample that is predominantly female. Most participants (31, 4 men) were from SK villages of the Middle Ucayali region (Yarinacocha, San Francisco, and Nueva Betania), with a subset from communities of the Lower (Paoyhan) and Upper (Puerto Belén) Ucayali region. Within the small town of Yarinacocha (in the vicinity of Pucallpa), we recruited participants (9, 2 men) from Bena Jema, a predominantly SK neighborhood. All the other recruitment sites were native community villages with exclusively SK residents but a strong relation with the outside world. Overall, the sample included SK adults who could be characterized as more urban (Yarinacocha and San Francisco sites) or more traditional and in regular contact with the surrounding rainforest (Nueva Betania, Paoyhan, and Puerto Belén sites).

The median age for participants was 38 years ($IQR = 26$-48) ranging from 20 to 64 years. Regarding occupations, 41% of the 32 female participants were homemakers or housewives (33% of the overall sample) and another 41% were artisans (33%). Three of the 7 male participants (43%, 8% overall) were horticulturalists. Across both sexes, 5 women (16%, 13% overall) and 3 men (43%, 8% overall) identified as students, comprising a total of 21% of the population. Although all adult participants were required to be native SK speakers, all were introduced to the Spanish language prior to adolescence (median age = 8yo, $IQR = 5$-10).

Materials and procedure. Similar to the original WCS, we used a set of 330 Munsell color chips and asked participants to name them (Berlin & Kay, 1969). We made a number of changes to the procedure, however. In the WCS, every participant provided terms for all 330 chips. Due to fear of participant fatigue, we split up color chips based on their ID numbers (even or odd) and participants were randomly assigned to work with either even- or odd-numbered color chips. As a result, each participant was presented with only 165 chips. All 330 hues within the set are visualized in Appendix 1. Dimensions of the chips were 2 cm × 2.5 cm.

First, the experimenter explained the general procedure and goals of the study to the participant. The experimenter would then present a single color chip to the participant and ask in SK: "What is the color of this chip?" The study was conducted solely in SK language with the assistance of a bilingual SK- and Spanish-speaking research assistant. It should be noted that although the experiment was conducted in SK, the SK word for color used is identical to the Spanish word *color* (an example of SK speakers adopting Spanish words into their lexicon), which might have encouraged Spanish language use.

Besides the reduction in set size, our procedure also differed from that of WCS (see Kay et al., 2009, pp. 585–591) in other aspects. Participants sat in front of the experimenter. To manage changes in natural light intensity between participants, the experiment took place indoors near a window or a door instead of outdoors. Another difference between our study and the WCS procedure is in our approach for encouraging participants to describe chips using BCTs. In the WCS, the experimenter would instruct participants to only provide BCTs during the task (e.g., describing a chip as "blue" as opposed to "navy blue" or "sky-like"). However, we had difficulties concisely explaining the concept of a BCT compared to other terms.[3] We decided to allow participants to describe a chip with any term they wished, and to ask further questions to elicit a BCT when they did not do so on their first try. For example, when presented with a red color chip, the participant might use the term "blood-like" (a non-BCT). The experimenter would then ask: "Do you know of any other word to refer to the color of this chip?" Should the participant subsequently respond with "dark red" (another non-BCT), the experimenter would further ask: "How would you refer to this color with only one word?" Eventually, the participant might use the term "red" (a BCT). For some chips, participants provided a BCT as their first description. For others, a BCT might be preceded by 1 or 2 non-BCTs. When participants failed to provide a BCT after 3 attempts (i.e., two follow-up questions), no further questions were asked, and the experimenter moved on to the next chip. All responses, BCT or not, were recorded in the order produced by the participant.

Results and Discussion

Figure 1 compares the original WCS data (Panel A) to a summary of results (Panel B) along with the prevalence of Spanish-language responses (Panel C) for Experiment 1. All participants used the following set of color terms to describe a color chip at least once during their session: "joxo" (light/white), "wiso" (dark/black), "panshin" (yellow), "joshin" (red), and "yankon" (green/blue). Given the widespread use of this term set and their interpretations, we will refer to these five terms as SK-language BCTs.

Most (79%) participants also described at least 1 chip as "manxan" (faded), referring to a chip's saturation. In addition, fifty-one percent of participants used the color term "naranja" (or "naransha") to describe at least one chip. Naranja may be known as a Spanish-language color term used to describe both the orange fruit and its associated color–as opposed to "anaranjado," a term strictly for the orange color.

In terms of overall popularity, participants described a median of 32% of chips as "yankon" ($IQR = 26$-39%) followed by "joshin" ($Mdn = 10\%$, $IQR = 7$-16%), "panshin" (10%, 6-12%), "joxo" (9%, 6-15%), "manxan" (6%, 1-10%), and "wiso" (5%, 3-8%). We failed to find

[3]Indeed, as Kay et al. (2009, pp. 587–589) acknowledge, there is no straightforward necessary and sufficient criteria for the basicness of a color term (cf Levinson, 2000).

any significant sex differences in the overall spread of color term usage across chip set ($t(59) = 0.00$, $p > .999$) or in the proportion of subjects who used a term at least once during their session ($t(117.95) = -0.38$, $p = .706$).

Compared to the WCS dataset which only reported SK language terms, Spanish use was prevalent throughout Study 1. Fifty-nine percent of our participants used a Spanish-language color term to describe at least 1 chip, which accounted for only 4% of all responses. Across chips, the most common Spanish-language color term was "naranja" (51% of participants), followed by "rosa" (10) and "morado" (8). Spanish use peaked at 16% when participants were asked to label chips that English speakers would consider to be orange or "anaranjado"/"naranja" by Spanish speakers. Indeed, the relatively common use of "naranja" by these adult SK speakers despite being prompted entirely in SK brings the possibility that "naranja" has been adopted into the SK color lexicon. If we allow "naranja" to be counted as an SK rather than Spanish-language term, then only 15% of participants used a Spanish-language term other than "naranja" at least once throughout the study, accounting for 2% of all responses. One participant responded in Spanish in 68% of the time despite being prompted solely in SK.

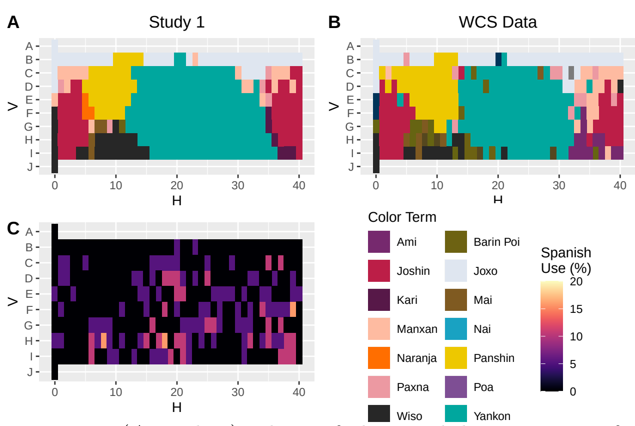

Figure 1. (A and B) Plots of the modal term given for a particular chip. Color coordinates were represented in 2-D Munsell space. Modal responses were given by SK adults during (A) our Study 1 and during (B) the original World Color Survey. (C) Heat map of prevalence of Spanish-language responses during Study 1. Legends for all three subplots located in the bottom-right quadrant.

Participants used an SK-language BCT (i.e., "yankon") to describe a median of 68% of chips (IQR = 56-90%). Besides BCTs, 59% of participants used SK-language ad hoc hue terms (i.e., "nai" or sky for blue chips) for an overall median of 6% of chips (IQR = 0-19%). SK-language terms referring to saturation or luminosity of a chip, such as "manxan" (faded) were used for an overall median of 13% of chips (IQR = 6-20%). Most instances (86%) of Spanish use involved a Spanish BCT such as "rojo" (overall Mdn = 0%, IQR = 0-0%).

Table 1
Demographics of participants in Studies 2 and 3.

Age Group	n	Boys
	Study 1	
5	3 (5% of overall sample)	1
6	8 (14%)	3
7	12 (21%)	4
8	15 (26%)	5
9	10 (18%)	5
10	4 (7%)	2
11	5 (9%)	3
	Study 2	
5	2 (4% of overall sample)	1
6	2 (4%)	0
7	11 (24%)	4
8	9 (20%)	1
9	11 (24%)	4
10	8 (17%)	3
11	3 (7%)	3

In sum, our data show similar variability to the WCS data, but with Spanish terms (as described above) mixed in with ad hoc terms. Notably, we observed the modal term for a few chips to be loanwords from Spanish, in some cases already established as part of the SK vocabulary (the last seems to be the case of "naranja," "orange" in English), suggesting some fairly extensive borrowing of Spanish words due to the close relation between both languages in the studied communities.

Study 2

After generating an updated SK color term map using the responses from adult participants in Study 1, we designed Study 2 to assess child participants' production and comprehension of SK color terms. Because we did not think that we could feasibly ask children across a range of ages about more than 100 color chips, we selected a subset of chips representing the prototypical instances for prominent SK terms from Study 1.

Methods

Participants. Fifty-seven children (23 boys) ages 5- to 11-years-old were recruited in predominantly SK neighborhoods in Yarinacocha (Nueva Era and Bena Jema) and in Bawanisho, a native community settled along the Ucayali River, more than 500 kilometers southeast of Pucallpa. Recruitment occurred either through direct contact with interested parents or through their local school. If recruited via school, consent for participation had to be given by both teacher and parent. Outside of the school environment, consent was given by the parent.

Materials and procedure. Based on the findings of Study 1, we chose 8 color chips from our original set of 330 to serve as prototypical instances of major SK color terms. These color chips were blue (WCS n°1), green (n°234), red (n°245), white (n°274), yellow (n°297), black (n°312), greenish-yellow (WCS n°320), and purple (WCS n°325). Study 2 was conducted entirely in SK and participants were explicitly instructed to give responses in SK as opposed to Spanish. In the production and comprehension tasks, children sat at a table across from the experimenter with color chips arranged between them. The production task was always performed before the comprehension task.

Production task. Similar to Study 1, the experimenter introduced the participant to the general procedure and the goals of the study. The experimenter would then ask: "What is the color of this chip?" As in Study 1, we used follow-up questions to elicit a BCT when the child's initial response was not a BCT. In a departure from Study 1, we were more explicit in soliciting an SK-language response. When a participant provided a Spanish-language term, the experimenter would record their response but further ask: "What is the name of this color in SK?" If a participant could not respond with an SK term, the experimenter would not ask further questions and would move forward to the next chip. As a result, some children only produced SK non-BCTs or Spanish-language terms for particular chips.

Comprehension task. The comprehension task had a notably different procedure compared to the preceding production task or that of Study 1. We tested the comprehension of 9 SK color terms. The choice of these terms was based on common responses given by adult participants in Study 1. The color term prompts included BCTs: "yankon" (green/blue), "joshin" (red), "panshin" (yellow), "joxo" (white/light), "wiso" (black/dark). We also included non-basic but prominent terms as prompts which were "nai" (sky or sky blue), and "barin poi" (greenish-yellow, meaning the Sun´s excrement, also used to refer to an alga) and two dyads of non-basic terms: "pei" (leaf) and "xo" (unripe) to represent the color green along with "ami" (a type of tree used to dye fabrics) "pua" (sachapapa, a tuber) to represent purple. Children sat at a table across from the experimenter with the 8 color chips of the production task displayed between them. The experimenter asked: "Can you give me the [*color term*] chip?" Participants chose one of the 8 chips and their response was recorded.

Our findings from Study 1 suggested that color terms varied in their degrees of specificity. For example, "wiso" best describes a narrow range of very dark to black. By contrast, "yankon" could encompass blue, green, greenish-yellow, and purple; "joshin" could describe red, purple, and orange; "pei" or "xo" could describe green or greenish-yellow. In cases where a term could apply to multiple chips (i.e., "yankon"), the chip selected first would be removed from the table, leaving 7 remaining chips. The experimenter would then ask: "Can you give me another [*color term*] chip?" The participant would then pick another one of the 7 chips, have their response recorded, and so on. We prompted participants 4 times for "yankon" and 2 times each for "joshin" and "pei"/"xo"; every other term only received a single prompt. Due to the inherent ambiguity in term-hue pairings, accuracy for a child participant was coded based on adult responses given during Study 1. If at least 15% of adult participants in Study 1 associated a chip with a particular term, we coded a similar term-chip pairing from a child participant as correct. Some trials could have multiple pairings; in those cases, accuracy was scored as an average, rather than as dichotomous. For instance, if a child correctly chose 3 out of 4 chips for the "yankon" trial, instead of 1 (correct) or 0 (incorrect) they would receive a score of 0.75.

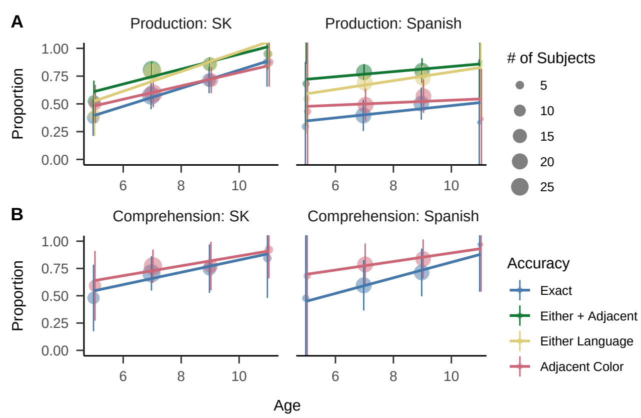

Figure 2. Proportion of accurate responses when applying different accuracy criteria, by age and study. Points show the mean for a 2-year age group (chosen arbitrarily for visualization) with 95% confidence intervals. Lines show a linear fit, weighted by the number of datapoints in each age group.

Results and Discussion

We begin by presenting general results from both the production and comprehension tasks, and then turn to specific analyses of overextensions. Figure 2 shows general trends across measures. For Study 2, we saw robust developmental increases in both production and comprehension. Because we had limited expectations regarding the amount of data that would be gathered during visits to the SK, we did not preregister our analyses. Thus all reported inferential statistics should be interpreted with some caution, and we do not adopt a specific cutoff of $\alpha = .05$ for interpretation.

To quantify these trends, we fit generalized linear mixed-effects models (GLMMs) predicting accuracy for both production and comprehension tasks with fixed effects of age in years (centered), random slopes of age for each color, and random intercepts for each color and participant. When these models failed to converge, we removed random slopes. We found highly significant age effects for both production ($\hat{\beta} = 0.85$, 95% CI $[0.46, 1.24]$, $z = 4.26$, $p < .001$) and comprehension ($\hat{\beta} = 0.40$, 95% CI $[0.18, 0.62]$, $z = 3.62$, $p < .001$). Most children in our study knew some SK color words, but few except some of the oldest children knew all of them.

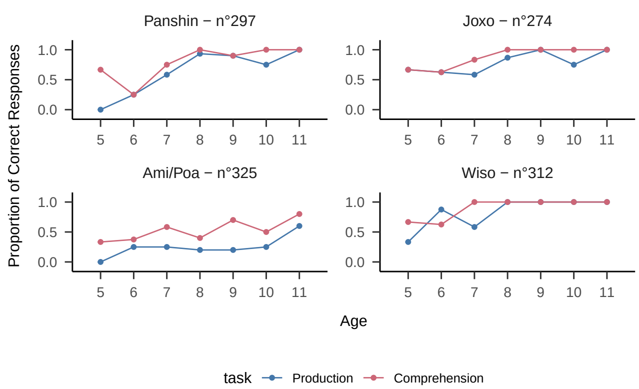

Figure 3. Production and comprehension data for selected color chips, plotted by age group.

Production vs. comprehension. While overall, production and comprehension accuracies were quite close, there were exceptions. For some term-chip pairings such as "ami/pua" and "pei/xo," children failed to produce the correct term in the production task but performed substantially better during the comprehension task (Figure 3). While there was a consistent ordering of tasks (production always first), there was no feedback on the production task, thus we think it is unlikely that children learned (or remembered) these labels as a function of task order. More likely is that these labels are relatively lower frequency and some children recognized them despite being unable to produce them.

Age of Acquisition. Following Michael C. Frank et al. (2021), we used the dichotomous responses given during the production task to predict the "age of acquisition" when at least half of SK children are predicted to properly label a particular chip. First, we split responses by the prompted chip for which each participant had a single entry. For each chip, we attempted to fit a generalized linear model by robust methods (Maechler et al., 2020) with the structure `accuracy of response [0 or 1] ~ age`. The coefficients for age ranged from 0.33 (odds of success multiplied by $\exp(\hat{\beta}) = 1.40$ with every added year of age) to 1.35 (odds multiplied by $\exp(\hat{\beta}) = 3.80$). To find age of acquisition then predicted the probability of success for the range of participant ages, 5.40- to 11.70-years-old at increments of 0.05 years, and selected the earliest age at which the accuracy crossed 0.5.

Using this method, we predict that half of SK children first learn to label the "joxo" chip (white) at 9.47.86.25.47.25.59.511.7 years of age. This is followed by the "wiso" chip (black) at 5.5, the "hoshin" chip (red) at 6.2, the "panshin" chip (yellow) at 7.2, the "yankon" chip (green) at 7.8, the "nai" chip (sky-blue; "yankon" also accepted) at 9.4, and the "yankon"/"joshin" chip (greenish-yellow) at 9.5. The model for one chip ("purple") did not predict that age of acquisition would have been met within our age range, with an estimated probability of 46% of children successfully labelling at 11.5 years of age.[4]

Our predictions suggest that SK children obtain color term knowledge at notably older ages compared to children in the United States (Wagner et al., 2013). Further, the ordering of acquisition is substantially different from that attested in previous studies. It is an interesting question what properties of children's input or the color terms themselves lead to this order of acquisition. Following Yurovsky et al. (2015), we might speculate about the potential that "joxo" is substantially higher frequency in SK than "white" is in English.

Language switching. Over a quarter (27%) of all responses were given in Spanish, despite children being prompted solely in SK (i.e., labeling a *panshin* chip as "amarillo"). The distribution of Spanish responses was non-random, with median use in 2/8 trials (IQR = 0-5). We did not find a significant correlation between

[4]It is worth noting that in Study 1, adult participants used 7 different labels for this chip (*ambi*, *ami*, *jimi*, *joshin*, *kari*, *morado*, and *yankon*), none of which were used more than 25% of the time.

Table 2
Naming entropy by color chip and whether the chip was used in Study 2 and Study 3.

Chip ID	Entropy	Study 2	Study 3	Shipibo term	Spanish term
1	0.71	×		Nai	Celeste
46	1.72		×	-	Gris
65	1.21		×	-	Rosa
121	1.49		×	-	Naranja
234	0.00	×	×	Pei/Xo	Verde
245	0.21	×	×	Joshin	Rojo
266	0.82		×	-	Marrón
274	0.33	×	×	Joxo	Blanco
291	0.90		×	-	Azul
297	0.21	×	×	Panshin	Amarillo
312	0.80	×	×	Wiso	Negro
320	1.34	×		Barin Poi	Medio rojo
325	1.94	×	×	Ami/Poa	Morado

age and number of trials with Spanish-language responses throughout the production task ($t(55) = -0.97$, $p = .335$).

As a further exploratory analysis, we attempted to quantify low naming consensus using naming entropy (following Gibson et al., 2017). We computed the naming entropy for each chip by computing the probabilities for each chip c to be named with a particular label l ($p(l \mid c)$) and then taking $H(c) = -\sum_l p(l \mid c) \log[p(l \mid c)]$ (see inset entropy values by chip in Table 2). To assess the hypothesis that naming entropy in adults was related to Spanish use in children, we fit a GLMM to predict likelihood of switching languages from SK to Spanish (a binary variable) as a function of child age, entropy of the chip's naming distribution for adults in Study 1, and their interaction (as well as random effects of subject). Despite age not being very correlated with overall frequency of Spanish responses, within this model, we found that overall older children tended to be less likely to respond in Spanish ($\hat{\beta} = -0.44$, 95% CI $[-0.96, 0.07]$, $z = -1.69$, $p = .092$), perhaps due to greater knowledge of SK terms. Children were also more likely to respond in Spanish when presented with a chip with high entropy (low naming consensus) among adult participants in Study 1 ($\hat{\beta} = 1.70$, 95% CI $[1.15, 2.24]$, $z = 6.10$, $p < .001$). We further found some evidence of a positive interaction between age and entropy ($\hat{\beta} = 0.30$, 95% CI $[-0.03, 0.62]$, $z = 1.78$, $p = .074$), suggesting more Spanish responding for older children specifically for those chips with high adult uncertainty. Together these findings suggest that older children show an increasingly adaptive use of Spanish vocabulary to describe chips for which there is not community consensus among adults.

Overextensions. One reason to use Spanish would be if children fail to recall the proper SK color term but do know the proper mapping in the Spanish. But another possibility is that children may have more imprecise representations and choose to respond with a same-language but adjacent color term (i.e., labeling a *pashin*-colored chip as "joshin"). Following Wagner et al. (2013), we aggregated across color chips and examined the pattern of children's first responses, categorizing them as same-language, adjacent, and different-language. We used a GLMM to assess whether calculated word entropy and age were associated with frequency of adjacent responses. We predicted the outcome using fixed effects of age in years (centered) and entropy, with random effects of participant. We found that younger children were more likely to respond with SK-language adjacent terms ($\hat{\beta} = -0.96$, 95% CI $[-1.58, -0.34]$, $z = -3.02$, $p = .002$) but chip entropy did not predict this strategy ($\hat{\beta} = -1.38$, 95% CI $[-3.06, 0.29]$, $z = -1.62$, $p = .106$). Further, coefficients in this model were almost identical to the coefficient for strict scoring, confirming the impression that these overextensions were relatively rare compared to the use of Spanish terms.

Study 3

Noting the apparent strategy of language switching seen in Study 2, we designed Study 3 as its complement. Here, we tested children's production and comprehension of Spanish color terms with a similar protocol to Study 2 but with a subset of chips representing prototypical colors for the Spanish color system.

Methods

Participants. We recruited a separate sample of 46 children (16 boys) ages 5- to 11-years-old from the neighborhood of Bena Jema in Yarinacocha and from Bawanisho. Recruitment occurred either through interested parents or a local school. As in Study 2, we received consent from parents and, if in a school environment, teachers as well.

Materials and procedure. Based on Study 1, we selected 11 color chips to serve as prototypical instances of prominent Peruvian Spanish color terms. These color chips included 6 also used during Study 2: green (n°234), red (n°245), white (n°274), yellow (n°297), black (n°312) , and purple (n°325). Five additional chips were selected: gray (WCS n°46), pink (n°65), orange (n°121), brown (n°266), and blue (n°291) (see Appendix 1). The blue chips differed between Studies 2 and 3 as we decided that the prototypical hues for *yankon* and *azul* differed enough to warrant the use of a different chip.

As we found that many SK children in our sample were not very fluent in Spanish – despite receiving some school instruction in Spanish – the production and comprehension tasks were both conducted in SK, and Spanish was only used for color terms (i.e., Spanish color terms were embedded within otherwise SK sentences). In both tasks, a participant would sit at a table across from the experimenter with 11 color chips in front. As in Study 2, the production task was always performed prior to the comprehension task.

Production task

The procedure was similar to that of both Studies 1 and 2. The experimenter would introduce a participant to the general procedure and aims of the study. Despite much of the study being conducted in SK, the experimenter would specify that participants would be expected to provide color terms in Spanish. The experimenter would then ask: "What is the color of this chip?" If the participant responded in SK, the experimenter would record their response but further ask: "What is the name of this color in Spanish?" If a participant responded with "I don't know" to this prompt, the experimenter would not prompt any further and would move forward to the next chip. As a result, some responses lack Spanish-language BCTs and only consist of non-basic and/or SK color terms. In total, we collected production data for 11 color chips. For each chip, the data include either one response (when children provided a Spanish basic color term in the first trial) or two or three responses (when children's initial responses were either non-basic and/or in SK).

Comprehension task. The procedure was similar to that of Study 2. The experimenter would ask: Can you give me the [*color term*] chip? For 11 Spanish color terms. The choice of these terms was based on both previous studies examining Spanish color terms as well as responses given by adult participants in Study 1 (as some adult participants used Spanish color terms to label particular color chips). The 11 terms used as prompts were "blanco" (white), "verde" (green), "rojo" (red), "amarillo" (yellow), "azul" (blue), "negro" ("black"), "naranja" (orange), "gris" (gray), "morado" (purple), "marrón" (brown), and "rosa" (pink). Since each color term was best instantiated by a single color chip and lacked the ambiguity seen with certain SK color terms, we defined a correct response as choosing the single color chip that matched the word, in contrast to Study 2.

Results and Discussion

As in Study 2, we observed age-related changes in color term accuracy for both production and comprehension. Aggregate results are visualized in 2. To assess these, we again fit GLMMs for both production and comprehension tasks with an identical structure to Study 2. Age was a significant predictor of accuracy in the comprehension task ($\hat{\beta} = 0.63$, 95% CI $[0.21, 1.06]$, $z = 2.90$, $p = .004$), but the age effect was weaker in the production task ($\hat{\beta} = 0.33$, 95% CI $[-0.06, 0.72]$, $z = 1.65$, $p = .098$, see Figure 2).

Similar to Study 2, over a quarter (30%) of all responses were given in SK, despite being prompted to respond in Spanish. There was significant variation in language-switching with some children responding solely in Spanish while others responded to upwards of 9/11 trials in SK (Mdn = 5 trials, IQR = 1.25-6). We found only a marginal correlation between age and accuracy ($t(44) = 1.91$, $p = .063$) and no significant correlation between age and language-switching ($t(44) = 0.44$, $p = .663$).

To assess to our hypothesis that older children would have more Spanish-language exposure and color term knowledge, we included age as a predictor in our GLMM assessing the effect of adult color naming entropy on likelihood to switch languages from Spanish to SK, similar to the one we fit for Study 2. This model did not show a significant interaction between age and adult color naming entropy ($\hat{\beta} = -0.27$, 95% CI $[-0.63, 0.09]$, $z = -1.49$, $p = .137$), however one without the interaction term did show an entropy effect ($\hat{\beta} = -1.49$, 95% CI $[-2.07, -0.92]$, $z = -5.10$, $p < .001$), tending to respond in SK for low-entropy items (those that were presumably more prototypical for the SK words). There was no significant effect of age ($\hat{\beta} = -0.02$, 95% CI $[-0.49, 0.45]$, $z = -0.08$, $p = .939$). Across studies, it appears that children preferred to respond in SK when presented with a chip for which adults had high consensus about the SK label, and in Spanish for low-consensus chips.

Similar to Study 2, we adopted alternative scoring to accommodate language-switching from Spanish to SK (different-language) and adjacent same-language responses. We used a GLMM identical to that of Study 2 in order to assess if changes in scoring criteria were associated with significant changes in task performance for production. Age was again a weaker predictor for production accuracy even with this more lenient scoring ($\hat{\beta} = 0.25$, 95% CI $[-0.07, 0.58]$, $z = 1.53$, $p = .126$), in concordance with earlier analyses. However, we did find that participants had higher aaccuracy when we includ-edSK responses ($\hat{\beta} = 1.76$, 95% CI $[1.43, 2.08]$, $z = 10.46$, $p < .001$) or adjacent same-language responses ($\hat{\beta} = 0.51$, 95% CI $[0.20, 0.81]$, $z = 3.27$, $p = .001$). In sum, we find frequent use of language switching in both Studies 2 and 3, but only Study 3 exhibits significant use of same-language but adjacent terms.

We speculate that early, informal Spanish language exposure can explain the discrepancies seen in Studies 2 and 3. With limited knowledge of Spanish color terms, children may spontaneously supplement their color term knowledge with Spanish terms during SK-language Study 2 but struggle to succeed in a more systematic evaluation in Study 3. More generally, we see children relying on a mixture of strategies to communicate colors even in the absence of mastery in either language.

General Discussion

In three studies, we mapped the color vocabulary of the Shipibo-Konibo (SK) language and used these data to study the development of color vocabulary in SK children growing up in a bilingual environment. This effort fills a gap in studies of color word development in non-WEIRD cultures and more generally parallels other efforts to use methods from language development to study populations that are under-represented in developmental science (Fortier et al., under review; e.g., Piantadosi et al., 2014).

With respect to the adult data, we found that the SK color vocabulary was relatively unchanged over the generations since the original WCS. Several interesting observations emerged, however. First, there was substantial use of Spanish terms, even though the task was conducted in SK, likely because the adults were recognizing focal colors for Spanish BCTs that have no parallel in SK (e.g., "naranja" for orange). This finding suggests an adaptive use of color vocabulary from both languages to succeed on the labeling task; future work will be required to understand whether such strategies are used in naturalistic communication as well. Second, we noted substantial use of ad hoc color terms (including luminance terms). These terms were used more often in SK than in Spanish, supporting the idea that Amazonian languages may make greater use of ad hoc color terms (at least in naming tasks) than Spanish speakers (e.g., Everett, 2005). Our data do not speak to whether this use is due to a desire to succeed on specific experimental tasks or whether it is comparable to use in naturalistic contexts. Nevertheless our findings are reminiscent of a suggestion by Levinson (2000) that even purported BCTs in Yélî Dnye did not fully span hue space and were often supplemented creatively with ad hoc terms.

When we turned to the children's data, we observed a much longer developmental trajectory for color than is observed in modern US English-learning children. As noted by H.Bornstein (1985), however, it is a very recent development that color terms are mastered as early as they are –z one hundred years ago, English-speaking US children's timeline of acquisition looked broadly similar to that observed in our study for SK children. We can only speculate as to the drivers of this historical change, but the industrialization hypothesis propounded by Gibson et al. (2017) appears to be a reasonable starting point. That is, industrialization allows for the production of identical objects that are usefully distinguished by color labeling. This communicative pressure can then lead to differentiation of color terms on a historical timescale and – relevant to our study here – is a likely driver of faster acquisition of color words by children.[5]

We did not find strong evidence for overextension in children's SK production or comprehension (with one or two exceptions), though there was somewhat more evidence for overextension in Spanish. This asymmetry might be due to less systematic or consistent exposure to Spanish vocabulary, but this explanation is merely speculative. We did, however, observe robust evidence for mixing and competition between the SK and Spanish color systems. Children differentially used Spanish terms in Study 2 when there was high uncertainty about the SK label for a particular color chip among adults. Similarly, they reached into their SK vocabulary in Study 3 when there was high consistency in SK labels among adults. These findings suggest that children were using their bilingual vocabulary adaptively to choose terms that are more likely to be interpreted correctly. Further, they suggest a potential route for functionally-driven language change, such that Spanish terms are borrowed – and perhaps eventually conventionalized – by children in cases where adult input data indicate uncertainty about the appropriate SK label.

Comprehension is generally thought to proceed production in language development generally (Eve V. Clark, 2009; Michael C. Frank et al., 2021) and in color word learning specifically (Wagner et al., 2018). In our data

[5] These speculations are informed by personal experience; the children of one author both learned their color terms in their second year through repeated practice with sets of manufactured plastic artifacts that varied only in hue, providing ideal teaching examples.

we did not observe large asymmetries between comprehension and production. Yet this result comes with caveats. First, comprehension and production tasks are by their nature different and have different demands (Sandhofer & Smith, 1999); this was especially true in our case given that the two tasks were sequenced and scored somewhat differently. Thus, we have chosen not to make quantitative comparisons between accuracies across these two tasks. Second, production and comprehension may be especially divergent for the youngest children, those who have the most difficulty with phonological encoding and the motoric aspects of production (Michael C. Frank et al., 2021); there is less evidence for production-comprehension divides in middle childhood. One natural question is whether comprehension and production dissociated in earlier times when US English-learners similarly acquired colors late; unfortunately we do not know of data that could be used to evaluate this question.

Our data here are consistent with models of color word meaning in which color word use is driven by functional need and languages adapt by developing vocabularies that appropriately allow for communication about those needs (Gibson et al., 2017; Zaslavsky et al., 2018). These models have not yet been generalized to either the bilingual setting or the acquisition setting, however. Our data suggest that functional language use can cross language boundaries, inviting models that consider code switching and borrowing as part of the process of change (e.g., Myslin & Levy, 2015).

Studying SK children's learning provides a descriptive comparison to studies of color naming in children learning English in the US (the focus of the majority of developmental work). Nonetheless, it has a number of limitations, some shared with this previous literature and some due to the specifics of our study and context. First, we regrettably do not have access to the kind of deep ethnographic observations that would allow us to hazard generalizations about how color terms are used in daily life among the SK communities we studied. Second, our study of development is cross-sectional and does not afford precision regarding the specific knowledge state of individual children due to the limited length of the task. Third, the limited number of color chips that we investigated means that our ability to generalize about the precision of particular color generalizations is much more limited for the children than the adults (limiting our entropy analyses). Finally, and perhaps most prominently, the kinds of tasks that we used are likely more unfamiliar to all of our participants and especially our child participants than they are to the populations being tested in investigations of WEIRD cultures (e.g., US English-learning children). While the performance of the oldest children in our studies was close to ceiling, the lower performance observed with younger children could in principle be in part a product of task unfamiliarity or other factors.

Going beyond convenience populations in experimental research with children is a new frontier for developmental science (Nielsen et al., 2017). Our work here suggests some of the benefits and challenges of this approach. On the positive, we can compare and generalize models of acquisition that are largely based on a single language and population (US English-acquiring children). At the same time, there is a paucity of resources describing language use, home environment, and cultural practices once we venture outside of WEIRD contexts. To best understand acquisition across cultures, we must document both children's knowledge and the structure of their environments.

References

Apicella, C. L., & Barrett, H. C. (2016). Cross-cultural evolutionary psychology. *Current Opinion in Psychology*, *7*, 92–97.

Aragón, K. (2016). *Color language and color categorization* (G. Paulsen, M. Uusküla, & J. Brindle, Eds.). Cambridge Scholars.

Bartlett, E. J. (1977). *Semantic organization and reference: Acquisition of two aspects of the meaning of color terms*. Biennial Meeting of the Society for Research in Child Development.

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

Bornstein, M. H., Kessen, W., & Weiskopf, S. (1976). Color vision and hue categorization in young human infants. *Journal of Experimental Psychology: Human Perception and Performance*, *2*(1), 115–129.

Chater, N., & Christiansen, M. H. (2010). Language acquisition meets language evolution. *Cognitive Science*, *34*(7), 1131–1157.

Clark, Eve V. (1973). *Cognitive development and acquisition of language* (pp. 65–110). Academic Press.

Clark, E. V. (1987). *Mechanisms of language acquisition* (B. MacWhinney, Ed.; pp. 1–33). Psychology Press.

Clark, Eve V. (2009). *First language acquisition*. Cambridge University Press.

Culbertson, J., Smolensky, P., & Legendre, G. (2012). Learning biases predict a word order universal. *Cognition*, *122*(3), 306–329.

Dobson, V., & Teller, D. Y. (1978). Visual acuity in human infants: A review and comparison of behavioral and electrophysiological studies. *Vision Research*, *18*(11), 1469–1483.

Everett, D. L. (2005). Cultural constraints on grammar and cognition in pirahã another look at the design features of human language. *Current Anthropology*, *46*(4), 621–646.

Fortier, M., Kellier, D., Fernández Flecha, M., & Frank, M. C. (under review). *Ad-hoc pragmatic implicatures among shipibo-konibo children in the peruvian amazon*.

Frank, Michael C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2021). *Variability and consistency in early language learning: The wordbank project*. MIT Press.

Frank, M. C., & Goodman, N. D. (2014). Inferring word meanings by assuming that speakers are informative. *Cognitive Psychology*, *75*, 80–96.

Franklin, A., Pilling, M., & Davies, I. (2005). The nature of infant color categorization: Evidence from eye movements on a target detection task. *Journal of Experimental Child Psychology*, *91*(3), 227–248.

Gibson, E., Futrell, R., Jara-Ettinger, J., Mahowald, K., Bergen, L., Ratnasingam, S., Gibson, M., Piantadosi, S. T., & Conway, B. (2017). Color naming across languages reflects color use. *Proceedings of the National Academy of Sciences*, *114*(40), 10785–10790.

H.Bornstein, M. (1985). On the development of color naming in young children: Data and theory. *Brain and Language*, *26*(1), 72–93.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*(2-3), 61–83.

Kay, P., Berlin, B., Maffin, L., Merrifield, W. R., & Cook, R. (2009). *The world color survey*. Center for the Study of Language; Information.

Kristol, A. M. (1980). Color systems in southern italy: A case of regression. *Language*, *56*(1), 137–145.

Lathrap, D. W. (1970). *The upper amazon*. Thames; Hudson.

Levinson, S. C. (2000). Yélî dnye and the theory of basic color terms. *Journal of Linguistic Anthropology*, *10*(1), 3–55.

Lillo, J., González-Perilli, F., Prado-León, L., Melnikova, A., Álvaro, L., Collado, J. A., & Moreira, H. (2018). Basic color terms (BCTs) and categories (BCCs) in three dialects of the spanish language: Interaction between cultural and universal factors. *Frontiers in Psychology*, *9*.

Maechler, M., Rousseeuw, P., Croux, C., Todorov, V., Ruckstuhl, A., Salibian-Barrera, M., Verbeke, T., Koller, M., Conceicao, E. L. T., & Anna di Palma, M. (2020). *Robustbase: Basic robust statistics* (R package version 0.93-6).

Monroy, M., & Custodio, S. (1989). Algunos usos de los terminos del color en el español de colombia. *Thesaurus; Bogotà*, *44*(2), 441.

Morin, E. (1973). *Le paradigme perdu: La nature humaine*. Éditions du Seuil.

Myers, T. P. (1974). Spanish contacts and social change on the ucayali river, peru. *Ethnohistory*, *21*(2), 135–137.

Myslin, M., & Levy, R. (2015). Code-switching and predictability of meaning in discourse. *Language*, *91*(4), 871–905.

Nielsen, M., Haun, D., Kärtner, J., & Legare, C. H. (2017). The persistent sampling bias in developmental psychology: A call to action. *Journal of Experimental Child Psychology*, *162*, 31–38.

Piantadosi, S. T., Jara-Ettinger, J., & Gibson, E. (2014). Children's learning of number words in an indigenous farming-foraging group. *Developmental Science*, *17*(4), 553–563.

Regier, T., Kay, P., & Khetarpal, N. (2007). Color naming reflects optimal partitions of color space. *Proceedings of the National Academy of Sciences*, *104*(4), 1436–1441.

Sandhofer, C. M., & Smith, L. B. (1999). Learning color words involves learning a system of mappings. *Developmental Psychology*, *35*(3), 668–679.

Sedivy, J. C. (2003). Pragmatic versus form-based accounts of referential contrast: Evidence for effects of informativity expectations. *Journal of Psycholinguistic Research*, *32*(1), 3–23.

St. Clair, K. (2016). *The secret lives of colour*. John Murray.

Surrallés, A. (2016). On contrastive perception and ineffability: Assessing sensory experience without colour terms in an amazonian society. *Journal of the Royal Anthropological Institute*, *22*(4), 962–979.

Tournon, J. (2002). *La merma magica: Vida e historia de los shipibo-conibo del ucayali*. Centro Amazonico de Antropologia Yaplicacion.

Wagner, K., Dobkins, K., & Barner, D. (2013). Slow mapping: Color word learning as a gradual inductive process. *Cognition*, *127*, 307–317.

Wagner, K., Jergens, J., & Barner, D. (2018). Partial color word comprehension precedes production. *Language Learning and Development*, *14*(4), 241–261.

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.*

Zaslavsky, N., Kemp, C., Tishby, N., & Regier, T. (2018). *Color naming reflects both perceptual structure and communicative need* (Vols. 1250–1255). Proceedings of the 40th Annual Conference of the Cognitive Science Society.