



OPEN ACCESS

EDITED BY

Ana Belén Barragán Martín,
University of Almería, Spain

REVIEWED BY

Siqi Zhang,
Stanford University, United States
Samuel Forbes,
Durham University, United Kingdom

*CORRESPONDENCE

Jiale Yang
✉ jyang@lets.chukyo-u.ac.jp

RECEIVED 05 June 2025

ACCEPTED 03 October 2025

PUBLISHED 27 October 2025

CITATION

Yang J, Dobkins K, Wagner K, Sakuta Y,
Kanazawa S and Yamaguchi MK (2025) Slow
mapping in color word acquisition across
languages: evidence from Japanese children.
Front. Dev. Psychol. 3:1641593.
doi: 10.3389/fdyps.2025.1641593

COPYRIGHT

© 2025 Yang, Dobkins, Wagner, Sakuta,
Kanazawa and Yamaguchi. This is an
open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](#). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic practice.
No use, distribution or reproduction is
permitted which does not comply with these
terms.

Slow mapping in color word acquisition across languages: evidence from Japanese children

Jiale Yang^{1*}, Karen Dobkins², Katie Wagner², Yuiko Sakuta³,
So Kanazawa⁴ and Masami K. Yamaguchi⁵

¹School of Psychology, Chukyo University, Nagoya, Japan, ²Department of Psychology, University of California, San Diego, La Jolla, CA, United States, ³Faculty of Human Life Sciences, Jissen Women's University, Tokyo, Japan, ⁴Department of Psychology, Japan Women's University, Tokyo, Japan, ⁵Department of Psychology, Chuo University, Tokyo, Japan

It is well-known that children have a delay between their first production of color words and acquisition of adult-like understanding. A previous study showed that this delay could be attributed to a process of gradually converging on language-specific color word boundaries. In this study, we tested this account in a second language, Japanese. We presented 12 color samples to children and then conducted production and comprehension tasks to check whether children have adult-like understanding of color words. Our results were consistent with previous findings showing that children before acquiring adult-like understanding tend to use color words systematically as overextensions of adult meanings. These results indicate that the delay between production and adult-like understanding of color words reflects a gradual process of learning language-specific color boundaries, potentially shared across languages.

KEYWORDS

categorization, color categories, word learning, perception, development

Introduction

For decades, color words have been a focus of word learning research due to the challenge they pose to children acquiring language. In the United States, children typically acquire color words around the age of 2–3 years of age (Shatz et al., 1996). However, there remains a significant delay between initial production and adult-like usage of color words (e.g., using “red” to exclusively label red objects). That is, by this age, children show a partial understanding of color terms, which is followed by a protracted period of refinement before converging on adult-like usage of color terms (Sandhofer and Smith, 1999; Pitchford and Mullen, 2002). This pattern seen for acquisition of color terms is mirrored in children's learning of other abstract domains, such as time (Tillman and Barner, 2015) and numerosity (for review, see Wagner et al., 2016).

Acquiring color words requires both identifying color properties as the correct domain of meaning and determining the boundaries between color word categories. In the domain of color categorization, several habituation studies suggest that infants exhibit early emerging color categories (between 4 and 6 months, Bornstein et al., 1976; Franklin and Davies, 2004; Skelton et al., 2017). More direct evidence for categorical color perception comes from studies showing that infants possess enhanced discrimination at category boundaries. For example, Franklin et al. (2008) used an eye-tracking visual search paradigm

and showed that infants (4.5 months) more rapidly detect a color target when it crosses a category boundary compared to when it remains within the same category. Providing a potential neural correlate for these perceptual findings, Yang et al. (2016), using near-infrared spectroscopy (NIRS) in young infants (between 5 and 7 months), found greater neural activation in response to between-category color changes than to within-category changes. Although the precise nature of categorical color perception in infancy remains open to debate, these findings collectively point to the emergence of categorical structure in infant color perception. Accordingly, most modern accounts of children's difficulty with color word learning assume that children can readily determine category boundaries by mapping newly acquired words onto pre-existing perceptual color categories.

Given the belief that color word boundaries may be pre-determined by perceptual categories, many explanations of color word acquisition attribute children's difficulty to a struggle with identifying color properties as the appropriate target domain for word meanings. These explanations argue that children are either unable to abstract the concept of color (Kowalski and Zimiles, 2006; Sandhofer and Smith, 1999) or simply do not attend to color properties because of word learning constraints (Franklin, 2006; O'Hanlon and Roberson, 2006; Pitchford and Mullen, 2002; Soja, 1994). Another widely discussed factor contributing to this difficulty is the whole-object bias (Macnamara, 1972; Markman, 1989), which refers to children's tendency to interpret new words as labels for whole objects rather than for object properties. In other words, the very mechanisms that may help children acquire labels for common objects may hinder their ability to acquire color words and other property labels.

While accounts that attribute the delay between color word production and adult-like usage of color words are parsimonious with both our understanding of the constraints that guide word learning and data demonstrating categorical perception of color in pre-linguistic infants, these accounts fail to provide an explanation of how children converge on language-specific color word meanings. Color terms vary widely across languages, not only in the number of basic color terms, but also in how perceptual space is divided (Berlin and Kay, 1969; Kay and Regier, 2003). Cross-cultural studies suggest that these linguistic differences shape perceptual categorization: both adults and children categorize colors in ways that reflect the structure of their native language (Roberson et al., 2000, 2004). As Maule et al. (2023) review, this cross-linguistic variation raises the possibility that the boundaries between colors are perceptually malleable and shaped by the linguistic categories a child learns. For instance, while English treats blue and light blue as a single category, Japanese uses two distinct terms: *ao* for blue and *mizuiro* for light blue. Such differences may influence how children form category boundaries and could impact the pace and pattern of color word acquisition. Supporting this idea, Saji et al. (2020) demonstrated that Japanese children's understanding of the term orange develops gradually through its contrast with other culturally defined color terms, suggesting that learning a color word involves understanding its place in the broader lexical system. Thus, while pre-linguistic perceptual color categories may serve to constrain the number of possible color word boundaries that children must consider, they cannot fully

define color word meanings. Any complete account of color word acquisition must provide a mechanism for how children converge on language-specific color word boundaries.

Building on this idea, Wagner et al. (2013) proposed a Slow Mapping hypothesis to explain the delay between production and adult-like usage of color words. They argue that the delay stems from the gradual process of refining boundaries between color word categories rather than difficulty identifying the correct domain of meaning for color words. In their study, children were presented twice with 11 colored objects and asked to label them. Results showed three key patterns: first, children's naming errors were consistent across the two instances (e.g., using *blue* to label for orange both times). Second, children's color words were often applied correctly (e.g., using *red* to label red) but overextended to *other* colors (e.g., using *red* to label orange or blue). Third, overextension errors were more likely to occur with adjacent color categories (e.g., *red* was more likely used to label orange vs. blue). Similar patterns were observed in a comprehension task, where children were asked to select a named color. This systematic nature of children's errors suggests that (1) children identify color properties as the correct domain of meaning very early in the acquisition process and (2) the delay between production and adult-like usage of color words reflects a gradual process of converging on language-specific color word boundaries.

The findings of Wagner et al. (2013) provide compelling evidence for a slow mapping process in English-acquiring children. However, languages differ widely in how they divide the perceptual color space and in the number and meaning of basic color terms. Thus, it remains an open question whether the same slow mapping processes apply cross-linguistically. To address this issue, the current study replicates Wagner et al.'s experimental design in Japanese, a language with a distinct color lexicon. This cross-linguistic replication allows us to test the generalizability of their findings and to explore whether the mechanisms underlying color word learning are shared across languages. Unlike Wagner et al. (2013), we tested both color word comprehension and production in the *same* children, enabling direct comparisons between these abilities. This within-subjects design increases statistical power, providing a more robust assessment of how children acquire and use color words.

Methods

Participants

Thirty-one Japanese children participated in the study. All children were raised in monolingual Japanese-speaking households. An a-priori power analysis was not conducted. Instead, the sample size was determined in line with Wagner et al. (2013), whose Experiment 2 involved 28 children in the word comprehension task. We therefore adopted a comparable sample size to ensure consistency with the precedent and comparability with the original study. This approach allows for comparability with the original study and supports direct replication. Furthermore, we later performed a *post hoc* sensitivity analysis (see Results section) to assess the statistical power achieved with our final

sample size. Two children were excluded: one for making no errors during any task and another for not cooperating. Data from the remaining 29 children (16 girls, 13 boys) were analyzed. Their ages ranged from 2.5 to 3.8 years (mean = 3.11; SD = 0.31). We chose this age range because it falls within a period in which slow mapping is refined, and partial comprehension of color terms has been reported from around 2 years of age (Wagner et al., 2013) and even earlier. Children were screened for protanopia or deuteranopia based on family history. This study was carried out in Japan. It was approved by the Ethics Committee of Chuo University and conducted in accordance with the ethical standards outlined in the Declaration of Helsinki. Informed consent was obtained from all parents or legal guardians of the participating children.

Stimuli

We used a set of 12 basic Japanese color terms, which largely overlapped with the categories used by Wagner et al. (2013), but included *mizuiro* (sky blue) as an additional category. The selected color terms were *aka* (red), *orenji* (orange), *chairo* (brown), *kiro* (yellow), *midori* (green), *mizuiro* (sky blue), *ao* (blue), *murasaki* (purple), *pinku* (pink), *shiro* (white), *haiiro* (gray), *kuro* (black). Stimuli were created using matte color chips from the Practical Color Co-ordinate System (PCCS), developed by the Japan Color Research Institute. To determine the specific color chip for each term, we conducted a preliminary study with five native Japanese-speaking adults. Participants were presented with the PCCS color chips and asked to select the chip they considered most representative of each color term. The most frequently selected chip for each term was then chosen as the stimulus. See Table 1 for the corresponding Munsell coordinates of each stimulus color. *Mizuiro* (sky blue) was included due to its high consensus among Japanese speakers, comparable to other basic color terms (Kuriki et al., 2017; Uchikawa and Boynton, 1987). All stimuli were presented on a black background, following the procedure of Wagner et al. (2013). The experimenters confirmed that the background did not reduce the perceptual discriminability of any color stimuli. For the Fish Production and Fish Comprehension tasks, the color chip was cut into fish shapes. For the Book Production task, the color chip was cut into squares and covered with white flaps of various shapes.

Procedure

The procedure was identical to that used in Wagner et al. (2013), including the tasks, materials, and instructions. All verbal instructions were translated into Japanese and adapted for naturalness, but no structural modifications were made.

Fish production task

Each child was presented with a black box containing the 12 colored fish, placed color-side down. The experimenter began by announcing, “Watashi no ban da! (My turn!)” and randomly

TABLE 1 Munsell coordinates for experimental stimuli.

Color	Munsell coordinates		
	Munsell hue	Value	Chroma
<i>aka</i> (red)	4R	4.5	14
<i>ao</i> (blue)	3PB	3.5	11.5
<i>chairo</i> (brown)	8YR	3.5	6.0
<i>haiiro</i> (gray)	N	5.5	–
<i>kiro</i> (yellow)	5Y	8.0	13.0
<i>kuro</i> (black)	N	1.5	–
<i>midori</i> (green)	3G	4.0	8.5
<i>mizuiro</i> (sky blue)	5B	7.0	5.0
<i>murasaki</i> (purple)	7P	2.5	9.5
<i>orenji</i> (orange)	4YR	6.0	14.0
<i>pinku</i> (pink)	6RP	7.0	6.0
<i>shiro</i> (white)	N	9.5	–

selected a fish, asking, “Kore wa nani iro? (What color is it?)” After the child responded, the experimenter placed the fish on the table and told the child, “Tsugi wa anata no ban da! (Your turn!),” indicating the child to pick up a fish and label it. This process continued alternately until all fish were labeled. If the child did not respond, the question was repeated, giving another chance to answer.

Book production task

Following the Fish Task, the experimenter presented a book containing 12 colored squares, each covered by a flap. For each page, the child lifted the flap, and the experimenter asked, “Kore wa nani iro? (What color is it?)” Colors were presented in a non-hue-based order: orange, blue, yellow, pink, white, purple, gray, brown, green, red, black, and sky blue. If the child did not respond, the question was repeated to provide another opportunity to answer.

Fish comprehension task

The 12 colored fish were placed color-side up in a random configuration. The experimenter asked the child to hand her a specific colored fish, saying, “(Akai) Sakana wo choudai. (Akai) sakana wo kudasai. [Give me a (red) fish. Give me a (red) fish.]” Once the child handed over a fish, it was returned to its place, and the next color was requested in random order. If the child did not respond (e.g., due to distraction), the question was repeated to give an additional opportunity to respond.

Results

In Table 2, we present raw performance data (averaged across 29 children) for each of the 11 colors. The first three columns

TABLE 2 Average performance ($n = 29$ children) on three trials: two production trials (fish and book) and one comprehension (fish) trial.

Color terms	Percent correct on each task			Percent of children Produced during production tasks
	Book production task	Fish production task	Fish comprehension task	
<i>aka</i> (red)	0.90	0.90	0.80	0.97
<i>ao</i> (blue)	0.76	0.90	0.79	0.93
<i>chairo</i> (brown)	0.38	0.50	0.69	0.52
<i>haiiro</i> (gray)	0.10	0.17	0.47	0.17
<i>kiiro</i> (yellow)	0.79	0.80	0.76	0.90
<i>kuro</i> (black)	0.79	0.86	0.93	0.90
<i>midori</i> (green)	0.72	0.76	0.69	0.76
<i>mizuuro</i> (sky blue)	0.45	0.34	0.50	0.55
<i>murasaki</i> (purple)	0.52	0.45	0.59	0.55
<i>orenji</i> (orange)	0.83	0.72	0.86	0.86
<i>pinku</i> (pink)	0.90	0.90	0.83	1.00

Note that the rules for determining whether a child is considered to have an “adult-like meaning” of a color term are explained in the text.

present mean percent correct data for each of the three trials (two production trials: fish vs. book, and the one comprehension trial). Also shown (far right column) for each color is the percentage of children who produced that color term at any point during the production trials. On average, children produced 9.34 (out of 11) color words during the production tasks ($sd = 2.27$; range = 4–12).

Acquisition of adult-like meanings

Children were classified as having an adult-like meaning for a color word if they (1) Correctly labeled the color on both production tasks (e.g., correctly labeled blue as *blue*), (2) Never provided the target label on any other production trials (e.g., never used *blue* for other stimuli), (3) Chose the correct stimulus in the comprehension task (e.g., correctly chose blue when asked for *blue*), and (4) Never selected the target stimulus for other colors in comprehension trials (e.g., never chose blue when asked for other colors). On average, children showed an adult-like meaning for 6.37 (out of 11) colors ($sd = 3.04$; range = 1–11).

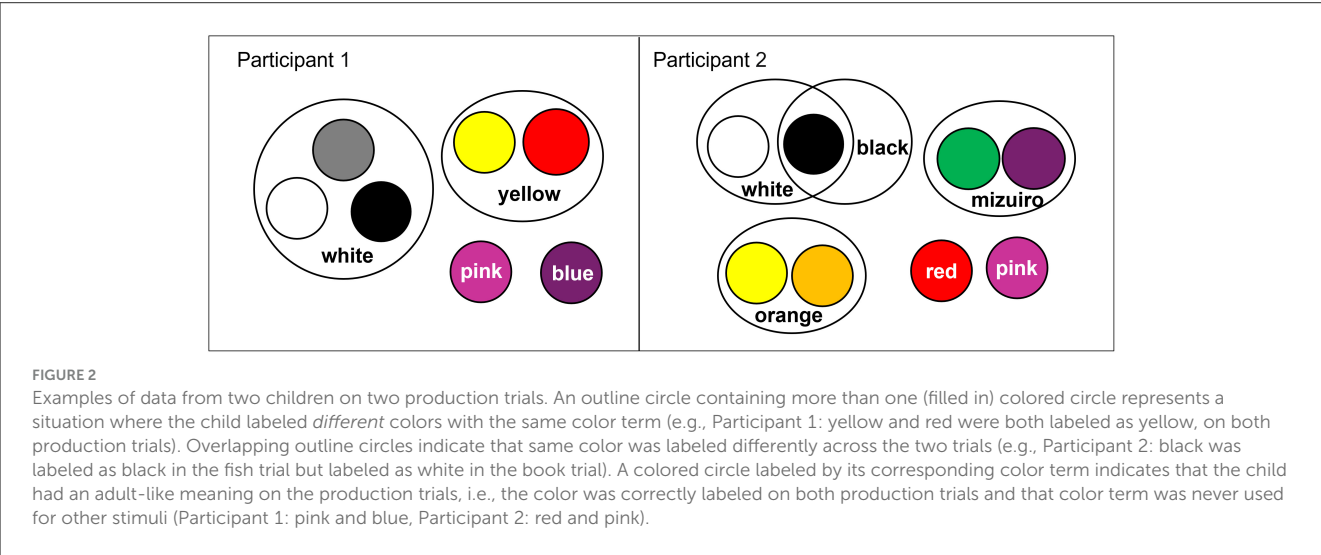
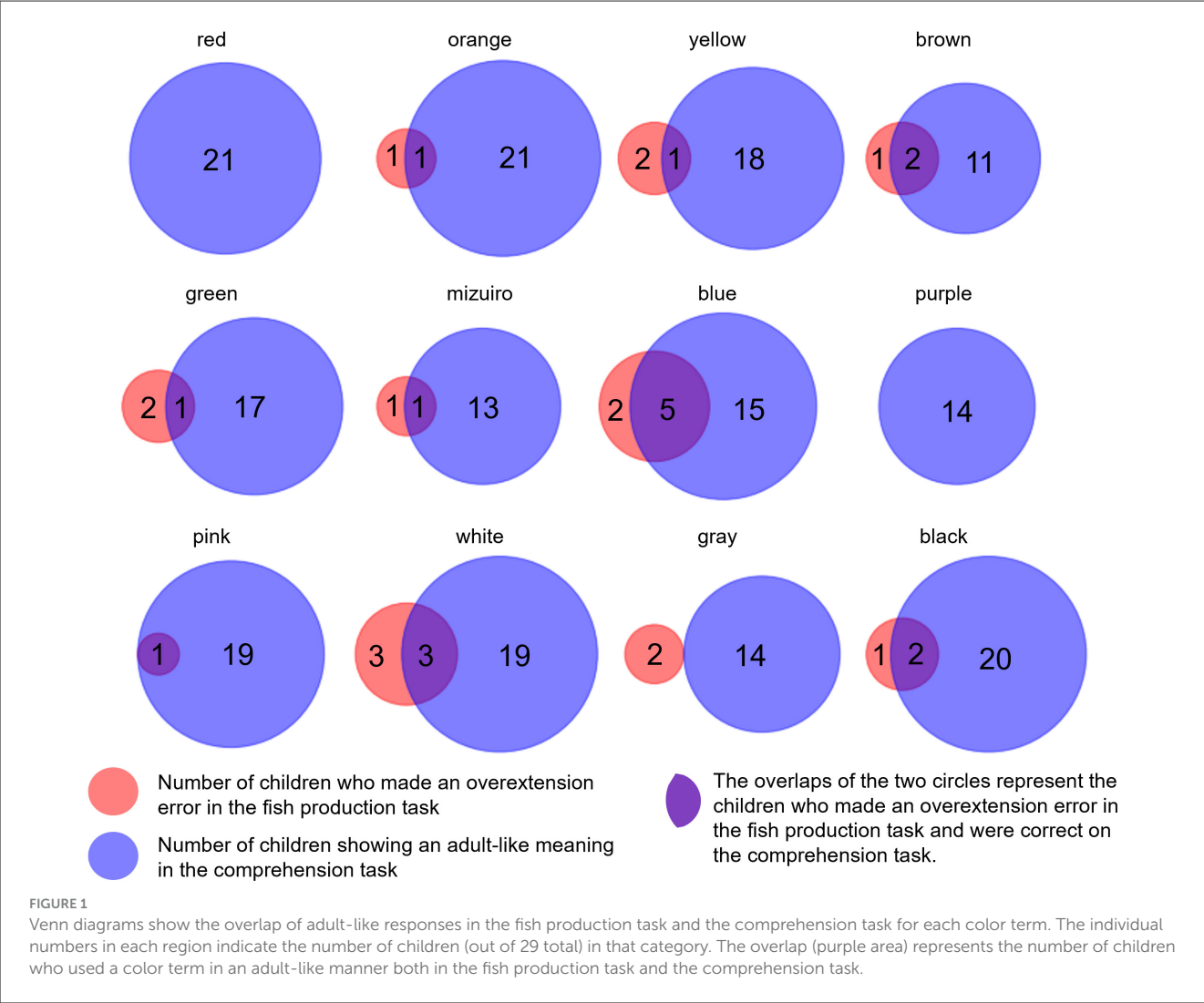
Figure 1 presents the number of children (out of 29) who exhibited adult-like meanings for each color in both production and comprehension tasks. Since each child participated in both tasks, we use a Venn diagram to illustrate the overlap. Note that for this analysis, we used only the results from the first production trial (i.e., the fish trial), as the comprehension task consisted of a single trial, making for a fairer between production and comprehension task performance. Importantly, analysis of error consistency in the following section revealed that children’s responses were significantly more consistent across the two production tasks than would be expected by chance. This suggests that the fish production task provides a representative sample of each child’s color word knowledge and is therefore appropriate for direct comparison with the comprehension task. Based on this comparison, the results reveal two key findings. First, there is substantial overlap between production and comprehension tasks (purple area), indicating that

children are consistent across tasks. Second, the extent of adult-like meanings varies by color: red had the highest proportion of children showing adult-like meaning (69%, 10 out of 29), while gray had the lowest (10%, 3 out of 29).

Error consistency analysis of production task

To analyze the consistency of errors across two production trials, we conducted an Error Consistency Analysis, in which we asked how likely it was that if the child made an error on one trial, they made that *same* error on the other trial. For example, as shown in Figure 2, Participant 1 consistently mislabeled red as *yellow* on both trials, whereas Participant 2 was inconsistent with black, labeling it incorrectly as white on one trial but correctly as black on the other. Using a binomial test, we examined whether the proportion of consistent trial pairs exceeded chance levels. Trial pairs where the child labeled the stimulus correctly on both tasks (216 pairs) were excluded. The remaining 132 pairs were classified as either consistent (same incorrect label on both trials, 62 pairs) or inconsistent (different incorrect labels or one correct and one incorrect label, 70 pairs).

To calculate chance-level consistency across the two production tasks, we adopted the method described by Wagner et al. (2013). For each child, we first calculated the base rate use of each color term (i.e., the proportion of total trials in which a given label was used) and squared this value to estimate the probability of repeating that label by chance. For example, if a child used the label “red” on 6 out of 22 trials, the probability of using “red” on both the fish and book trials by chance would be $(6/22)^2$. We then summed these squared probabilities across all labels used by the child to obtain their overall chance-level consistency. Finally, group-level chance consistency was computed as a weighted average of individual children’s values, with weights reflecting the number of relevant



stimulus pairs contributed by each child. In other words:

$$p(\text{consistency}) = \sum_c \sum_j \left(\frac{l_j}{n}\right)^2 \left(\frac{i_c}{i}\right)$$

where i is the total number of stimulus pairs in which at least one label (either book or fish) was incorrect, i_c is the number of such incorrect pairs that each child, c , contributed to the analysis, l_j is the number of times a child produced each label j and n is the total number of responses a child produced.

The proportion of consistent incorrect trial pairs (47%) was significantly greater than chance (17%) in a binomial test ($p < 0.001$, Cohen's $h = 0.66$). A *post hoc* power analysis using the pwr package in R (version 4.2.2) indicated high power (97.2%). This indicates that the sample size was adequate to detect the observed effect. This result indicates that children's color labeling errors were highly consistent, despite differences in stimulus shapes between the fish and book trials. This supports the idea that children develop interim meanings before attaining adult-like understanding of color words. However, the Consistency Analysis alone does not reveal the nature of these errors. It is possible that children make errors in a systematic way. For example, a child might correctly label yellow but also overextend it to green and orange. To explore this, we conducted two additional analyses.

Overextension analysis of production task

To test whether children's color errors were overextensions of adult color categories, we conducted an Overextension Analysis. An error was defined as an overextension if the child used a given color word correctly for its target color on both the fish and book production tasks and also used that same word incorrectly for a non-target color. For example, if a child incorrectly labeled orange and yellow as *red*, we checked whether they also labeled red as *red*. Based on these criteria, Participant 1's use of *yellow* for red counted as overextension (see Figure 2), since he also labeled *yellow* correctly for yellow. In contrast, Participant 2's use of *mizuiro* for green and blue was not, because she did not use *mizuiro* for sky blue. Note that consistency across multiple incorrect uses was not required; rather, the label needed to be used consistently for its correct referent across both tasks.

To estimate the probability of overextension errors occurring by chance, we adopted the method described by Wagner et al. (2013). For each child, we first identified all color terms that were incorrectly used (i.e., produced for a non-target color) and calculated the base rate of each term, defined as the proportion of total production trials in which the child used that label. We then squared the base rate for each incorrectly applied label to estimate the probability that the same incorrect label would appear on both trials featuring the correct color stimulus (e.g., using "red" on both the red fish and red book trials). These probabilities were then averaged across all such instances for each child, yielding an individual-level chance probability of overextension. Finally, a group-level estimate was computed as the weighted mean of the individual probabilities, with weights corresponding to the number of incorrect labels each child contributed to the analysis. Based on

this procedure, the overall probability of overextension errors was calculated as:

$$p(\text{overextension}) = \sum_c \frac{i_c}{i} \frac{\sum_j \left(\frac{i_{cj}}{n}\right)^2}{i_c}$$

where i is the total number of labels that were used incorrectly at least once, i_c is the number of such incorrect labels that each child, c , contributed to the analysis, i_{cj} is the number of times a child produced each incorrect label j , and n is the total number of responses a child produced.

Using a binomial test, we asked whether the proportion of overextension errors exceeded chance. The participants produced 46 color words incorrectly. To assess whether these errors reflected overextensions, we identified all cases in which a color word was used incorrectly by a participant. This yielded 46 unique color words that had been used incorrectly. However, to qualify as an overextension, the same word must also have been used correctly for its target color in both the Fish and Book Production Tasks. Therefore, 5 instances were excluded because the relevant color label was not produced for the target color in either production task, making it impossible to determine whether the label was ever used correctly. Among the remaining 41 errors, 78% were overextensions, closely matching Wagner et al. (2013) (76%). This rate was significantly above chance (3.6%; $p < 0.001$, Cohen's $h = 1.78$; *post hoc* power = 1.00), and comparable to the chance level reported by Wagner et al. (2013), which was 5.4%. This low probability reflects children's sparse and uneven use of color words, which results in low base rates for many labels. Following Wagner et al. (2013), the chance level of overextension rate was estimated by first calculating the base rate of each color word separately for each child, squaring these values, and then computing a weighted average across children. This method accounts for individual variability in label use and provides a conservative estimate. To count as an overextension, a label must be used correctly for the target stimulus on both production tasks. For reference, if all 12 color words were used equally, the chance of a child correctly using the same label for the target stimulus in both production tasks would be $1/12^2 \approx 0.7\%$. This illustrates how sparse and consistent label use leads to low chance expectations, making our estimated chance level of 3.6% a conservative benchmark. The fact that 78% of errors were overextensions suggests that when children produced a color word, they were very likely to use it consistently and accurately for its target color. It suggests that most errors were overextensions anchored to adult-like focal hues (i.e., hues that are the best examples of adults' categories).

Proximity analysis of the production task

The overextension analysis indicates that children not only use color words for their target colors but also extend them to label other colors before acquiring adult-like color word meanings. If overextensions were random, one would expect children to use incorrect labels indiscriminately across the color space. However, if the delay between production and adult-like usage of color

words reflects a gradual process of converging on language-specific color word boundaries, then overextension errors should be more likely to occur between perceptually similar (i.e., proximal) colors. Proximity was defined based on perceptual similarity in the Munsell Color System, where adjacent categories represent proximal colors. For example, red is closely related to orange, pink, purple, and brown, while blue and red are considered non-proximal. For more details, refer to [Wagner et al. \(2013\)](#). A child might label green as *blue* rather than *yellow*, reflecting the greater similarity between green and blue. To test this prediction, we conducted a proximity analysis, examining whether incorrect labels were from perceptually proximal color categories. As in [Wagner et al. \(2013\)](#), this analysis included all incorrect labeled stimulus pairings, regardless of the consistency of label use. For example, Participant 1 in [Figure 2](#) made a proximal error by labeling gray as *white*, whereas labeling red as *yellow* was a non-proximal error (as red and yellow are not proximal categories). Using a binomial test, we asked whether the proportion of proximal errors exceeded chance. Chance probability of proximal errors was determined based on the frequency of errors for each stimulus and the frequency of incorrect label use. Specifically, the chance of a given label-stimulus pair was calculated as the product of the base rates for that label and that stimulus. For example, if 20% of errors involved the orange stimulus and 80% involved the red label, the chance of labeling orange as red was estimated as $0.2 \times 0.8 = 0.16$. The overall chance probability was calculated by summing the probabilities of all label-stimulus pairs as proximal. The overall probability of proximal errors was calculated as:

$$p(\text{proximity}) = \sum_i \sum_j p(r | l_j \cap s_i) p(l_j | \text{incorrect}) p(s_i | \text{incorrect})$$

where, $p(s_i | \text{incorrect})$ is the probability of a particular stimulus i given an incorrect response; $p(l_j | \text{incorrect})$ is the probability of a particular elicited label j given an incorrect response to stimulus i ; and r is the probability of proximity. Note that $p(r | l_j \cap s_i)$ is either 1 or 0 because a given label/stimulus pair is either proximal or not proximal.

Production performance on the book trial (percent correct = 67%) and fish trial (68%) were not significantly different [$t_{(30)} = 0.48$, $p = 0.60$]. Therefore, data from both trials were combined for analysis. Among all errors, 66% involved proximal categories, which was significantly higher than chance (31%; $p < 0.001$, Cohen's $h = 0.71$; *post hoc* power = 1.00). This indicates that the errors made by children were likely to be labels for perceptually similar colors.

Proximity analysis of comprehension task

The purpose of the comprehension task, which consisted of a single trial per color word, was to investigate the same “proximity” question as conducted above for the production task. However, overextension could not be analyzed in the comprehension task because it only allowed one response per item. This design prevents us from determining whether children associate a given color word

with both its correct referent and other non-target colors, a key requirement for identifying overextensions. Six participants were excluded from this analysis, as they made errors in production but not in comprehension tasks. Among 261 trials from the remaining 23 participants, 82 trials (31%) were errors and were included in the analysis.

As in the production tasks, we accounted for the base rate of errors involving each color stimulus. However, on the comprehension task, our calculation of chance accounted for the base rate of errors that involved each stimulus and the base rate of errors made to a particular request (e.g., *red*) rather than using the base rate of errors that involved each stimulus and the base rates of incorrect labels produced by the child. This change reflects a structural difference between the tasks: whereas the production task involved child-generated labels, the comprehension task required children to choose a color patch in response to a verbal label. Thus, we based our estimate of chance on the joint probability of selecting a given stimulus in response to a given request. Aside from this difference in how chance was estimated, the analytical procedure followed the same steps as in the proximity analysis of the production task. Consistent with the results from the production task, 36% of the errors in comprehension task involved proximal colors, which was significantly above chance (27%; $p < 0.037$, Cohen's $h = 0.19$; *post hoc* power = 0.42).

To examine whether proximal error rates differed between comprehension and production, we directly compared the proximal error rate of these two tasks. The proximal error rate was defined as the proportion of errors that involved perceptually proximal color categories, relative to the total number of errors in each task. Because the rate is undefined when there are no errors, we included only participants who made at least one error in both tasks ($n = 20$). A paired-samples t -test revealed a significant difference between tasks [$t_{(19)} = 2.394$, $p < 0.027$, Cohen's $d = 0.54$; *post hoc* power = 0.63], indicating that proximity errors comprised a higher proportion of errors in the production task than in the comprehension task.

Discussion

The Slow Mapping hypothesis ([Wagner et al., 2013](#)) attributes the delay between color word production and acquisition of adult-like meanings to the gradual process of refining boundaries for color words. While the original results were based on American English-speaking children, our study strengthens the hypothesis by replicating these findings in Japanese-speaking children. In the production task, we examined children's color word production errors. Consistent with [Wagner et al. \(2013\)](#), our findings reveal that 3-year-old children systematically assign meanings color words. First, in the Error Consistency Analysis, children were highly consistent in their color words usage across the fish and book production trials, even when their usage differed from adult-like meaning. This indicates they can abstract color across objects with varying features and shapes, suggesting that even early in acquisition, children form stable, albeit non-adult, hypotheses about color word meanings. Second, in the Overextension Analysis, children correctly labeled focal colors (e.g., *red* for a red stimulus) but often overextended the same label to other colors (e.g., orange

and yellow) in both production and comprehension tasks. This indicates that children understand the focal meaning of color words but extend their usage to include other colors, suggesting that children understand the core meaning of color words and tend to use them accurately for focal colors, even if they overextend them to other hues. Thirdly, in the Proximity Analysis, children frequently made proximal errors, using the same color word for perceptually adjacent colors. This finding suggests that children initially form broad, overextended linguistic color categories before acquiring more precise, adult-like meanings. This pattern supports the idea that children's early categories are shaped by perceptual structure but require refinement through language-specific input. Taken together, these results suggest that (1) 3-year-old children can abstract color across objects early in color word acquisition, even without adult-like usage, and (2) they pass through a stage with broad linguistic color categories, gradually narrowing these boundaries through an inductive learning process as they encounter more examples of word usage.

To our knowledge, this is the first study to demonstrate that the Slow Mapping process (Wagner et al., 2013) may be shared across linguistically and culturally distinct populations. Our findings align with evidence that even 3-year-old Japanese children can map color words to its typical referents (Saji et al., 2020). While some accounts argue that children easily map newly acquired color words to pre-existing perceptual categories (Shatz et al., 1996; Pitchford and Mullen, 2002), our results challenge this view. Specifically, we show that children abstract color early in word acquisition but often use color words within broad, overextended categories. Importantly, these errors are not random; they are consistent across tasks (e.g., fish and book production trials) and often involve proximal colors. For example, a child might use the word red not only for red but also for orange and yellow. This pattern highlights systematic processes underlying their early use of color words. Our findings suggest that the delay in achieving adult-like meanings arises not from difficulty with initial abstraction but from the gradual mapping of linguistic color category boundaries. The replication of these patterns in Japanese-speaking children supports the universality of broad, early color categories in language acquisition.

This interpretation is consistent with the foundational work by Carey and Bartlett (1978), who introduced the concept of fast mapping to describe children's ability to rapidly associate a novel label with a referent. However, they emphasized that fast mapping marks only the beginning of word learning, which is followed by a longer phase of refinement. According to this view, even after children begin to use color terms, their understanding remains incomplete and continues to refine over time. This view was supported by Bartlett (1978), who showed that children often made proximity-based errors and required several months to transition from partial understanding to full adult-like usage of color words. Our results build on this early work, demonstrating that similar overextensions are evident in Japanese, suggesting a shared, gradual process of semantic refinement across languages.

One notable aspect of this study is that both the production and comprehension tasks were conducted with the same children, enabling a direct comparison of their performance across tasks. To ensure comparability across studies, we intentionally adopted a design where the comprehension task included only one trial

compared to two trials in the production task, replicating the methodology of Wagner et al. (2013). As shown in Figure 1, the Venn diagrams reveal a substantial overlap between children who demonstrated adult-like meanings in the production task and those who did so in the comprehension task. This overlap indicates that children who correctly produce a color word are also likely to understand its meaning in comprehension, reflecting systematic patterns in their language use, regardless of whether children are asked to produce vs. comprehend a color.

The proximity analysis revealed that proximal errors were more frequent in the production task (66% of errors) compared to the comprehension task (36%). This difference may be partially explained by a communication strategy (Clark, 1978), wherein children possess adult-like meanings for color words but use them to label unknown colors. For example, a child who knows blue but not purple might label a purple object as blue, as blue is perceptually proximal and preferable to giving no answer. This suggests that, in production, children may rely on familiar color terms to approximate unfamiliar ones when faced with uncertainty. However, the significant proximal errors observed in the comprehension task, which does not require verbal production, suggest that the communication strategy alone cannot fully explain these findings. If children relied solely on a communicative strategy to avoid producing no response in the production task, we would expect the proportion of proximal errors in comprehension to be at chance. The observed proportion (0.36) was modest but significantly above chance (0.27). There are several reasons why we would not expect the rates of proximal errors to be very high. The main is that the comprehension task data includes responses to both words that a child truly does not yet know (these responses are expected to be at chance) and words for which a child has an emerging meaning. In the latter case, only a portion of the errors would be expected to be proximal. Consider a child that has an overextended category of *red* that includes the colors red, orange, yellow and green. This child, when asked for *red*, may respond by providing the red stimulus (correct response), the orange stimulus (proximal error) or a yellow or green stimulus (non-proximal errors). Despite a cohesive representation of the word *red*, only a portion of this child's errors would be predicted to be proximal.

One noteworthy difference between the present study and Wagner et al. (2013) is the inclusion of *mizuiro* (sky blue) as a distinct color category in Japanese. Unlike English, which typically treats light blue as a subcategory of blue, Japanese clearly distinguishes between *ao* (blue) and *mizuiro* (light blue) in adult usage (Kuriki et al., 2017; Uchikawa and Boynton, 1987). Support for treating *mizuiro* as a basic color term for this age group comes from children's actual performance. As shown in Table 2, children's accuracy in both comprehension and production tasks for *mizuiro* was comparable to that for other Japanese basic color terms such as *murasaki* (purple) and *chairo* (brown). Moreover, more than half of the children (55%) spontaneously produced the word during production tasks, further supporting its status as an age-appropriate and linguistically relevant category. This linguistic distinction underscores how color word acquisition is influenced not only by perceptual mechanisms but also by the structure of the lexicon in the child's native language.

Regardless of the exact nature of the proximal errors observed in the current study, these errors are likely to be linguistic rather than perceptual. This claim is supported by studies showing that pre-linguistic infants exhibit sharp perceptual boundaries between colors (Bornstein et al., 1976; Franklin and Davies, 2004; Skelton et al., 2017; Yang et al., 2016). Thus, the broadness of early color categories shown in the present study likely arises from linguistic, not perceptual, factors. This supports the idea that linguistic color boundaries develop by adapting to language-specific categories, which vary across languages. For instance, Russian distinguishes light blue (*goluboy*) from dark blue (*siniy*), and Korean includes *cheongnok* (turquoise) as distinct from green and blue. These cross-linguistic differences suggest that the Slow Mapping process could play a role in helping learners adapt perceptual categories to language-specific boundaries.

One methodological limitation of the present study is the fixed task order, with the comprehension task always administered after both production tasks. Although this decision defies conventional practice of counterbalancing, this decision was made to limit the extent to which the tasks can influence one another. During the production task, the child only hears color words that they are able to generate themselves in response to a question from the experimenter, *what color is this?* In contrast, on the comprehension task, the experimenter provides each of the 12 color words, i.e., *Give me the red fish*. Because of this asymmetry in the tasks, we believed it was more likely for the production task to influence the comprehension task than vice versa.

Additionally, although our study did not examine age-based differences in color word knowledge, this was also true of Wagner et al. (2013), who instead grouped children according to their productive color vocabulary. Such an approach offers valuable insight but typically requires larger sample sizes than were feasible in our study. Nevertheless, prior research suggests that children begin to use color words around age 2–3, with more consistent and adult-like usage emerging between ages 4 and 5 (e.g., Pitchford and Mullen, 2002; Sandhofer and Smith, 1999; Shatz et al., 1996).

Another limitation concerns the statistical power of some secondary analyses. While the primary findings of the study, such as consistency, overextension and proximal errors, were supported by large effect sizes and high statistical power (often exceeding 0.90), several exploratory comparisons (e.g., between proximal error rate in production and comprehension tasks) showed moderate effect sizes but relatively low power. These results, although statistically significant, should be interpreted with caution. They highlight promising directions for future research, which would benefit from larger samples to enable more robust tests of secondary effects and interactions.

To conclude, this study supports the Slow Mapping hypothesis in non-English-speaking children. The delay between color word production and mastery of adult-like meanings appears to result from a gradual process of converging color word boundaries, rather than an inability to identify the correct domain of meaning.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by Ethics Committee of Chuo University, Tokyo, Japan. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation in this study was provided by the participants' legal guardians/next of kin.

Author contributions

JY: Conceptualization, Data curation, Funding acquisition, Validation, Visualization, Writing – original draft, Writing – review & editing. KD: Conceptualization, Methodology, Supervision, Writing – review & editing. KW: Conceptualization, Data curation, Formal analysis, Methodology, Project administration, Validation, Visualization, Writing – original draft, Writing – review & editing. YS: Data curation, Investigation, Methodology, Project administration, Writing – review & editing. SK: Funding acquisition, Supervision, Writing – review & editing. MKY: Funding acquisition, Project administration, Supervision, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was supported by Grants-in-Aid for Scientific Research from JSPS (24H00702 to JY, SK, and MKY, and 21243041 to SK and MKY), and a Grant-in-Aid for Scientific Research on Innovative Areas “Shitsukan” from MEXT, Japan to MKY (25135729).

Acknowledgments

We sincerely thank the children and their parents for their participation and collaboration in this study.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Bartlett, E. J. (1978). "The acquisition of the meaning of color terms: a study of lexical development," in *Recent Advances in the Psychology of Language*, eds. R. N. Campbell and P. T. Smith (New York, NY: Plenum Press), 89–108.
- Berlin, B., and Kay, P. (1969). *Basic Color Terms: Their Universality and Evolution*. Berkeley, CA: University of California Press.
- Bornstein, M. H., Kessen, W., and Weiskopf, S. (1976). Color vision and hue categorization in young human infants. *J. Exp. Psychol. Hum. Percept. Perform.* 2, 115–129. doi: 10.1037//0096-1523.2.1.115
- Carey, S., and Bartlett, E. (1978). Acquiring a single new word. *Papers Rep. Child Lang. Dev.* 15, 17–29.
- Clark, E. V. (1978). Strategies for communicating. *Child Dev.* 49, 953–959. doi: 10.2307/1128734
- Franklin, A. (2006). Constraints on children's color term acquisition. *J. Exp. Child Psychol.* 94, 322–327. doi: 10.1016/j.jecp.2006.02.003
- Franklin, A., and Davies, I. R. L. (2004). New evidence for infant colour categories. *Br. J. Dev. Psychol.* 22, 349–377. doi: 10.1348/0261510041552738
- Franklin, A., Drivonikou, G. V., Bevis, L., Davies, I. R. L., Kay, P., and Regier, T. (2008). Categorical perception of color is lateralized to the right hemisphere in infants, but to the left hemisphere in adults. *Proc. Natl. Acad. Sci. U.S.A.* 105, 3221–3225. doi: 10.1073/pnas.0712286105
- Kay, P., and Regier, T. (2003). Resolving the question of color naming universals. *Proc. Natl. Acad. Sci. U.S.A.* 100, 9085–9089. doi: 10.1073/pnas.1532837100
- Kowalski, K., and Zimiles, H. (2006). The relation between children's conceptual functioning with color and color term acquisition. *J. Exp. Child Psychol.* 94, 301–321. doi: 10.1016/j.jecp.2005.12.001
- Kuriki, I., Lange, R., Muto, Y., Brown, A. M., Fukuda, K., Tokunaga, R., et al. (2017). The modern Japanese color lexicon. *J. Vis.* 17:1. doi: 10.1167/17.3.1
- Macnamara, J. (1972). Cognitive basis of language learning in infants. *Psychol. Rev.* 79, 1–13. doi: 10.1037/h0031901
- Markman, E. M. (1989). *Categorization and Naming in Children: Problems of Induction*. Cambridge, MA: MIT Press.
- Maule, J., Skelton, A. E., and Franklin, A. (2023). The development of color perception and cognition. *Annu. Rev. Psychol.* 74, 87–111. doi: 10.1146/annurev-psych-032720-040512
- O'Hanlon, C. G., and Roberson, D. (2006). Learning in context: Linguistic and attentional constraints on children's color term learning. *J. Exp. Child Psychol.* 94, 275–300. doi: 10.1016/j.jecp.2005.11.007
- Pitchford, N. J., and Mullen, K. T. (2002). Is the acquisition of basic-colour terms in young children constrained? *Perception* 31, 1349–1370. doi: 10.1068/p3405
- Roberson, D., Davidoff, J., and Davies, I. (2000). Color categories are not universal: Replications and new evidence from a stone-age culture. *J. Exp. Psychol. General* 129, 369–398. doi: 10.1037/0096-3445.129.3.369
- Roberson, D., Davies, I. R. L., Davidoff, J., and Shapiro, L. R. (2004). The development of color categories in two languages: a longitudinal study. *J. Exp. Psychol. General* 133, 554–571. doi: 10.1037/0096-3445.133.4.554
- Saji, N., Imai, M., and Asano, M. (2020). Acquisition of the meaning of the word orange requires understanding of the meanings of red, pink, and purple: constructing a lexicon as a connected system. *Cogn. Sci.* 44:e12813. doi: 10.1111/cogs.12813
- Sandhofer, C. M., and Smith, L. B. (1999). Learning color words involves learning a system of mappings. *Dev. Psychol.* 35, 668–679. doi: 10.1037/0012-1649.35.3.668
- Shatz, M., Behrend, D., Gelman, S. A., and Ebeling, K. S. (1996). Colour term knowledge in two-year-olds: evidence for early competence. *J. Child Lang.* 23, 177–199. doi: 10.1017/S030500090001014X
- Skelton, A. E., Catchpole, G., Abbott, J. T., Bosten, J. M., and Franklin, A. (2017). Biological origins of color categorization. *Proc. Nat. Acad. Sci. U.S.A.* 114, 5545–5550. doi: 10.1073/pnas.1612881114
- Soja, N. N. (1994). Young children's concept of color and its relation to the acquisition of color words. *Child Dev.* 65, 918–937. doi: 10.2307/1131428
- Tillman, K. A., and Barner, D. (2015). Learning the language of time: children's acquisition of duration words. *Cogn. Psychol.* 78, 57–77. doi: 10.1016/j.cogpsych.2015.03.001
- Uchikawa, K., and Boynton, R. M. (1987). Categorical color perception of Japanese observers: comparison with that of Americans. *Vision Res.* 27, 1825–1833. doi: 10.1016/0042-6989(87)90111-8
- Wagner, K., Dobkins, K., and Barner, D. (2013). Slow mapping: color word learning as a gradual inductive process. *Cognition* 127, 307–317. doi: 10.1016/j.cognition.2013.01.010
- Wagner, K., Tillman, K., and Barner, D. (2016). Inferring number, time, and color concepts from core knowledge and linguistic structure. *Core Knowl. Concept. Change* 105, 105–126. doi: 10.1093/acprof:oso/9780190467630.003.0007
- Yang, J., Kanazawa, S., Yamaguchi, M. K., and Kuriki, I. (2016). Cortical response to categorical color perception in infants investigated by near-infrared spectroscopy. *Proc. Nat. Acad. Sci. U.S.A.* 113, 2370–2375. doi: 10.1073/pnas.1512044113