1   Developmental changes in the precision of visual concept knowledge

Abstract

How precise are the visual representations that underlie children's understanding of words, and how does this precision change across development? We assessed children's visual vocabulary knowledge in a game where children heard a word (e.g., "swordfish") and had to choose a matching picture. Distractors were chosen to vary in their linguistic similarity to each target word in a multimodal language model. We collected data from a large sample of children (N = 3575, 3–14 year olds) and adults (N = 211 adults) and found gradual changes over development in children's ability to identify the image that a word referred to. When children made errors, younger children were more likely to choose unrelated distractors than older children; children's error patterns were best explained by combining linguistic and visual similarity. These results highlight that children have partial knowledge about many visual concepts and document a transition from coarse to finer-grained representations.

Keywords: visual concepts, receptive vocabulary, large language models, object recognition

Word count: 1999 excluding Method/Results

Developmental changes in the precision of visual concept knowledge

Introduction

Across the world's languages, many words refer to visible things in the environment (e.g. "thermos", "prism", "whale"). How precise are the visual representations that underlie children's understanding of these words—children's visual concepts? For example, depending on how old a child is (and how much they know about whales) a child might understand that the word whale refers to blue whales—but not whale sharks—or instead only know that a whale is a large animal that lives in the ocean. Children's acquisition of the boundaries of visual categories and how they map onto words is a key developmental process, as children who lag behind their peers in vocabulary learning are at increased risk for language delays with significant academic and social consequences (Rescorla & Dale, 2013).

Most empirical work on the precision of children's visual concepts has focused on the first few years of life, when children learn words at an astonishing rate (Bloom, 2000; Braginsky et al., 2019). In looking-while-listening tasks (Fernald et al., 2008), infants as young as 6 months look to events associated with common nouns like "bottle" and "banana" (Bergelson & Aslin, 2017), and 14- to 18-month-olds extend newly learned words to atypical exemplars of these categories (Weaver, Zettersten, & Saffran, 2024). By around their second birthday, children also generalize nouns to stylized, 3D exemplars (Pereira & Smith, 2009) as they learn that shape is a valuable cue to basic-level category membership. In parallel, empirical work on children's developing ability to recognize objects (Ayzenberg & Behrmann, 2024) has focused on these first few years of childhood as an important period of developmental change.

An implicit assumption underlying much of this research is that preschoolers and school-aged children's visual concepts—e.g., their knowledge of what whales "look like" —may not change substantially after they can understand the word "whale" in conversations. On this account, children associate words (e.g., "whale") with high-fidelity visual information

43 about a particular exemplar (or a few) during learning, and this is sufficient for acquiring the

44 visual concept. Instead, the enrichment for this concept is thought to happen solely at the

45 level of semantics (e.g., "whales eat krill"); on this semantic enrichment hypothesis, later

46 learning during the elementary school years does not influence the content of the visual

47 concepts per se.

48   But these assumptions may be incorrect. Perhaps these visual concepts themselves

49 undergo considerable and gradual change beyond the first few years of childhood. On this

50 alternative view, in parallel with learning an increasingly rich semantic space of word

51 meanings, children also develop an increasingly detailed representation of the diagnostic

52 features of these visual concepts, and a better understanding of the boundaries of these

53 visual categories. Indeed, classic theories have long articulated a view of this type, in which

54 children might initially set up a place-holder concept via "fast mapping" but then gradually

55 fill in later details (Carey & Bartlett, 1978; cf. Swingley, 2010; Dale, 1965). An example from

56 this early work focused on a case where three-year-olds remembered that a word

57 ("chromium") was a color word, but they did not know which color it was. Similarly, in

58 foundational theories of vocabulary acquisition, children's vocabulary knowledge has been

59 thought of as progressing along stages (Dale, 1965; Beck et al., 1987), with an intermediate

60 stage where children were thought to have a "vague knowledge of the word's meaning" (Dale,

61 1965) before they fully grasp what a word means. For example, there might be an

62 intermediate phase in which children understand that "whales" are animals that live in the

63 ocean, but have less precision on the visual features that distinguish them from fish and

64 other sea mammals.

65   Yet despite this long-standing recognition that children likely have partial word

66 knowledge (Dale, 1965; Carey & Bartlett, 1978), all assessments of receptive vocabulary

67 characterize word knowledge in an all-or-nothing fashion, perhaps because it is indeed

68 difficult to quantify what it would mean to have a low-precision representation for a word

69 (Dale, 1965; Carey & Bartlett, 1978). Both direct measures, such as the Peabody Picture

70 Vocabulary Test (Dunn et al., 2003; Gershon et al., 2014), and parent-report measures of

71 vocabulary knowledge (Braginsky et al., 2019) characterize children as either "knowing" a

72 word or not. Since the distractors for these assessments are chosen carefully, children's error

73 patterns could in principle reveal evidence for partial knowledge (e.g., selecting a whale

74 shark when the target word is whale), but the datasets resulting from these proprietary

75 assessments are unavailable for re-analysis.

76 In parallel, a large literature on expertise suggests that visual learning almost certainly

77 extends throughout early and middle childhood and into adulthood (for reviews, see Curby

78 & Gauthier, 2005; Harel, 2016). Decades of empirical work has established that birding

79 experts, car aficionados (Tanaka & Gauthier, 1997), and graphic artists (Perdreau &

80 Cavanagh, 2013), among others, have both qualitatively and quantitatively different kinds of

81 visual representations for the visual concepts that they have significant expertise with that is

82 reflected in both behavioral and neural responses. Thus, on a visual expertise hypothesis,

83 children's visual concepts might continually be enriched throughout early and middle

84 childhood. More recent developmental work supports this idea (Long et al, 2024): In a large

85 observational study, children became increasingly able to both depict and recognize line

86 drawings of visual concepts (e.g., "whale", "clock", "tiger") from 3-10 years of age, hinting at

87 underlying changes in how children represent the diagnostic features of these visual concepts.

88 Our theoretical accounts of visual concept development have thus been limited by (1) a

89 need for larger developmental datasets with dense data on individual items and (2) the

90 difficulty of systematically quantifying what it means to have a partial or imprecise

91 representation of a visual concept. Here, we take steps to address both of these challenges.

92 We created a new, open-source picture matching task and developed research partnerships

93 with schools across the country in order to gather dense item data across a wide age range

94 (3–14 years) from many children (N=3575 children, N=211 adults). We built a new,

open-source "visual vocabulary" task where children heard a word (e.g., "swordfish") and had to choose a picture in order to teach aliens about their meanings. We then collected data from a large sample of representative children in both online contexts as well as through research-practice partnership with preschools and elementary school contexts across the U.S. Next, we leveraged modern computational models to choose distractor items systematically and to analyze children's resulting responses. For each word we chose distractor items that varied in their linguistic similarity to the target word as operationalized via the similarity of the target and distractor words in the language encoder of a multimodal language model (Contrastive Language-Image Pre-training model, or CLIP; Radford et al., 2021), and we used embeddings from the visual encoder to model children's error patterns. These advances in data and models thus allow us to quantify changes in the precision of children's visual concepts across early and middle childhood for the first time.

With these data, we investigated three main questions. First, we examined whether we see evidence for continual learning of visual vocabulary knowledge throughout middle childhood (consistent with a visual expertise hypothesis) versus a plateau in performance in the elementary school years (consistent with a semantic enrichment hypothesis). Second, we examined evidence for changes in the precision of visual concepts before word knowledge is solidified, as per Dale (1965)'s account. If younger children's representations are initially overly broad, then we would expect younger children to select distractors of all types when they do not know a word's exact meaning, and older children to almost exclusively choose high-similarity distractors (e.g., whale sharks); if children bypass this stage where they have lossy representations, then children should choose distractors randomly. Finally, we examined children's error patterns: if changes in children's error patterns are driven by their knowledge of what visual information "goes with the word", then a combination of linguistic and visual similarity metrics should explain variance in children's error patterns. In contrast, if children are primarily drawn to visually-similar distractors when they do not know a word's meaning—irrespective of their linguistic similarity to the target word—then visual

similarity alone should be a driving factor in children's error patterns.

Method

.

Participants

To obtain a large sample of responses for individual items across development, we collected data from children across several different testing contexts, totaling $N = 3786$ unique participants. We collected data from children in an in-person preschool ($N = 65$, 3-5 year-olds), from the Children Helping Science (CHS) Platform, ($N=243$, 3-7 year-olds), and representative samples from schools across the United States through the [BLINDED] platform $N=3267$, 5-14 year-olds) (BLINDED et al.). We also recruited adults online ($N=211$) via Prolific; we sampled from monolinguals and adults who spoke English as a second language to maximize our chance of finding item variability. Schools included seven public school districts (14 schools), three independent schools specializing for students with learning differences, two independent schools, and one summer program; these diverse student data were obtained through research-practice partnerships. Most school-aged participants responded directly via clicking through a laptop or tapping through a tablet or a tablet, except those recruited online through CHS and Prolific; children's parents responded via clicking on the image on CHS, and adults responded via clicking on the images.

Children completed, on average, 25.02 trials that were sampled randomly from the stimulus set (maximum = 86; different maximum numbers of trials were included in different testing contexts). All participants who contributed data and scored above chance (chance = 25%, threshold of >30% accuracy) were included, regardless of the number of trials they completed (minimum trials = 2, maximum trials = 86. Some participants participated in the game more than once ($N = 214$ children), and their data are included as if they completed additional trials; note that our statistical models include random intercepts for each

participant and operate at the trial level; however, excluding these participants did not

change the pattern of results. We excluded an additional 84 participants who scored near

chance on 4AFC trials ($< 30\%$ accurate) and were school-aged (>6 years of age); these

participants completed an average of 17.72 trials.

Materials

We capitalized on publicly available existing image and audio databases to generate

trials. Visual concepts were taken from the THINGS+ dataset (Stoinski et al., 2023), after

filtering out non-child safe images (e.g., weapons, cigarettes) and images with

low-nameability ($<.3$), as per Stoinski et al., 2023. We used the copyright-free, high-quality

image released for each visual concept. We subset to visual concepts that had available

age-of-acquisition (AoA) ratings from a previous existing dataset (Kuperman et al., 2012)

where adults were asked to retrospect on the age that they learned a word in development;

these estimated AoAs were used as a proxy for item difficulty. Audio for each target word

was taken from the MALD database (Tucker et al., 2019).

We used these stimuli to create trials that contained a target image paired with

distractor images of similar difficulty (estimated using AoA) and varying similarity

(estimated using a computational model). To do so, we sampled distractors with high,

medium, and low similarity to the target word as operationalized via the embedding

similarity of the words in the language encoder of a multimodal large language model

(Radford et al., 2021).We determined high-, medium-, and low-similarity values relative to

the distribution of all possible target-distractor similarity values for each word in the

THINGS+ dataset. Stimuli were selected to optimize for having a maximum number of trials

with unique target and distractors; in addition, we constrained the sampling such that

target-distractor pairs had estimated AoA within 3 years of each other. For each target word,

we first selected a high-similarity distractor that had the highest cosine similarity to the

target (and was itself not one of the target words). For low-similarity words, we sampled a

unique distractor word that had the lowest cosine similarity among the remaining distractors. For medium-similarity distractors, we then randomly sampled a distractor word that was the same animacy as the target word, and unique to the dataset. In our final set, we had 108 items with a range of different estimated age-of-acquisitions (e.g., hedgehog, mandolin, mulch, swordfish, waterwheel, bobsled) with all unique targets and distractors. All stimuli and their metadata are available on the public repository for this project (BLINDED).

Model features

We obtained all model features using the Open AI available implementation of CLIP available at https://github.com/openai/CLIP. For language similarity, we computed the cosine similarity of the embeddings of the target word to each distractor word on each trial (e.g., tulip – rose, tulip – glove, tulip – hubcap). For visual similarity, we repeated this procedure but by obtaining image similarity vectors in the vision transformer for each target image and distractor image on each trial. For multimodal similarity, we computed the cosine similarity of the embedding of the target word in the language model to the embeddings for each of the distractor images; this is possible because the embedding spaces for the vision and language transformers in the CLIP model are aligned and have the same number of dimensions. All code is available on the repository associated with this project.

Procedure Children were invited to participate in a picture matching game where they were asked to "teach aliens on another planet about some of the words on our planet"; children picked a particular alien to "accompany them on their journey." Before the images appeared, children heard a target word (e.g., "apple") and then were asked to "choose the picture that goes with the word". The four images appeared in randomized locations on the screen, and one of the images always corresponded to the target word (see 1). On practice trials, the distractor images were all very dissimilar to the target concept, and the target word was relatively easy. The game played a chime sound if children chose the correct image, and a slightly unpleasant sound if they responded incorrectly; this feedback ensured that

children of all ages understood the task correctly. Each child viewed a random subset of the item bank, and the items they viewed were displayed in a random order. Children were allowed to stop the game whenever they wanted to. Different versions of the game included varying amounts of total possible trials and additional items, as these games were developed as part of an additional project to create an open-sourced measure of children's vocabulary knowledge (Blinded et al., in prep). Here, we analyze children's responses to items that were generated using the THINGS+ dataset with distractors of varying difficulty (see Stimuli).

Data and code sharing

These data were collected with the secondary goal of creating an open-source measure of children's developing vocabulary knowledge. These analyses presented in the paper were not pre-registered, and children's data were analyzed both during and at the end of data collection. All pre-processed data and analysis and plotting code have been made available in an online repository. The raw, trial-by-trial data cannot be shared due to data sharing agreements between the participating schools and the institution's IRB.

<div align="center">

Results

</div>

A protracted developmental trajectory

Do children gradually acquire visual vocabulary knowledge, or are precise representations for most nouns acquired by middle childhood? We found a gradual increase in children's ability to correctly identify the target picture across our entire age range; Figure 2 shows the proportion of time that children identified the target, highlighting a protracted developmental trajectory and little evidence for a plateau in the elementary years. Children became increasingly accurate over development at identifying the correct referent that a word referred to, with the oldest children in our sample (13- and 14-year-olds) still performing less accurately than adults.

We found a slight developmental trend for relatively easy words that had an average

224 estimated age-of-acquisition (Kuperman et al., 2012) (AoA) of 4.81 years (SD = 0.87; acorn,

225 bulldozer, scoop, puddle, sprinkler), but change over development was much more

226 pronounced for more difficult words (barrel, buffet, coaster, stump, carousel; average AoA =

227 6.95 years, SD = 0.65) and challenging words (e.g., tapestry, scaffolding, mandolin, aloe;

228 average AoA = 9.60 years, SD = 1.21) (see Figure 2). At an item level, 3-year-olds'

229 performance on the easiest words (watermelon, M = .93; turtle, M = .93, lollipop, M = .95)

230 suggests that they understood the task. However, other items with relatively low AoAs

231 (squirrel, M = .78, hamster, M = .72; map, M = .73) were more difficult for young children.

232 Over development, the words that showed the greatest change across age (see Figure 3a),

233 included some animals (e.g., swordfish), as well as inanimate objects (prism, antenna,

234 sandbag, turbine) but also other kinds of objects, including parts of buildings (scaffolding,

235 gutter). Some items, however, showed relatively little change across age: these tended to be

236 easy words that children across all ages identified correctly (sunflower, rice, potato,

237 hedgehog). Overall, these results highlight variability across individual items and suggest

238 that children gradually build knowledge of many visual concepts throughout childhood.

Changes in the precision of visual concepts

240   Do children begin with high-fidelity representations of what word referents look like or

241 are they acquired more slowly? Thus far, these analyses support a view where children

242 progressively acquire visual vocabulary knowledge via precise visual concepts (e.g., that can

243 distinguish between a whale and a whale shark) on a relatively long developmental timeline.

244 However, they do not yet establish whether children have visual concepts that become more

245 precise en route to mature word understanding. If children's visual concepts indeed have an

246 intermediate stage where they are overly broad (e.g., encompass both whales and whale

247 sharks), then we would expect younger children to be more likely to choose distractors of all

248 types (showing evidence for a very coarse representation or no representation at all), whereas

249 we would expect older children to almost exclusively choose related distractors (showing

evidence of partial knowledge). Alternatively, if children quickly acquire relatively

high-fidelity representations, then we should only observe changes in how accurate children

are at identifying the target word, with no changes in the types of distractors that children

choose when they choose incorrectly.

We examined whether there are systematic changes in how children made errors across

development, shown in Figure 3b. Consistent with the partial knowledge hypothesis, we

found that older children chose more closely related distractors than younger children.

Adults were still more likely to choose the related distractors than to the oldest children in

our sample (14-year-olds). We confirmed this result via a linear regression, modeling the

proportion of errors that each child chose related distractors as our dependent variable as a

function of children's age (in years); we also included the number of errors each child made

as a covariate as this varied widely by participant and age group. We found a main effect of

age (B = 0.04, SE = 0.01, t = 7.45, p < .001): older children were more likely to choose

linguistically related vs. unrelated distractors. However, we did not find strong differences

between distractors that were less related to the target word: both low-similarity and

medium-similarity distractors were equally unlikely to be chosen, suggesting that even

though model similarities distinguished these options, children did not. Overall, these results

suggest a progression where children gradually build up detailed knowledge about the visual

referents of many challenging words.

Modeling changes in children's error patterns
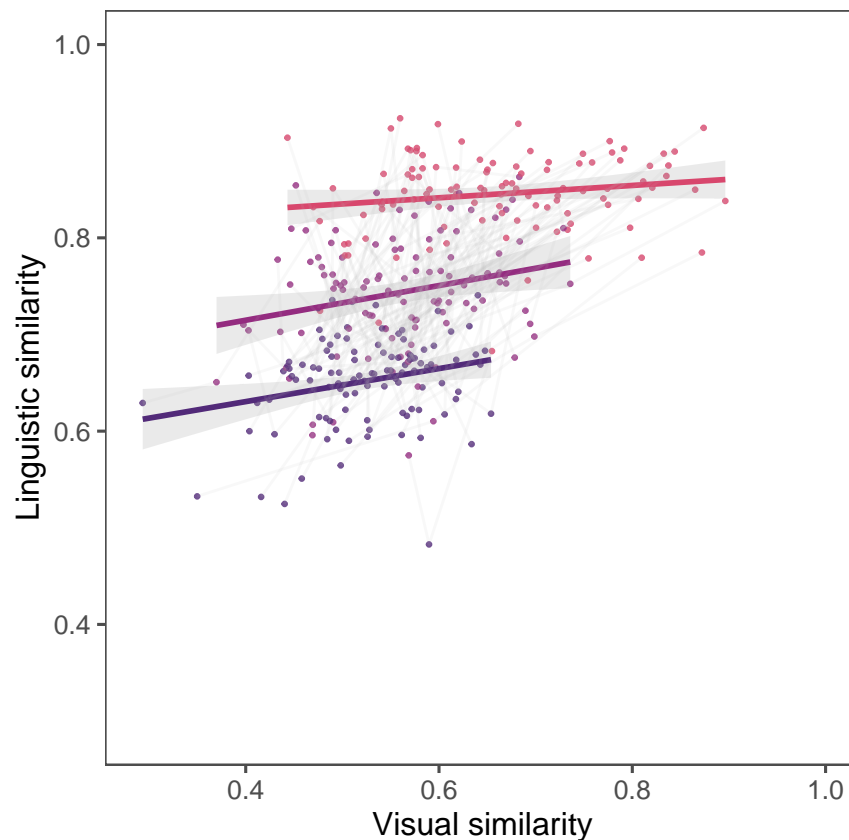
In a final set of analyses, we aimed to understand the source of changes in children's error patterns by leveraging similarity metrics derived from a large, multimodal language model (Radford et al., 2021). When we created our stimuli, we paired targets and distractors by using word similarity in a linguistic embedding space, but of course the corresponding visual stimuli naturally vary in their visual similarity to each other. This is true both at a category level— that is, animals are more similar visually to each other than to inanimate objects (Long et al., 2017; Krigeskorte et al., 2008) but also can be true at an image level—e.g., the flowers in the images of rose and tulip were both pink. Thus, some stimuli on certain trials could be more related to the target concept semantically and less related visually, or vice versa. For example, the words scoop and sauce had high similarity in the language encoder (r = .924), and their corresponding images had lower similarity in the vision encoder (r = .560). Conversely, the words tulip and rose had relatively lower similarity

in the vision encoder (r = .785) than their corresponding images in the vision encoder (r = .872) (see Figure 4A, Figures 4B). Thus, our large and varied set of stimuli—and the ability to easily parameterize the relative linguistic and visual similarity of all target-distractor pairs—affords the opportunity to examine what drives item-level variability in this task. We thus examined the degree to which children's error patterns reflected changes in how they processed the visual similarity of the targets and distractors, their semantic similarity, or—perhaps most likely—some combination. To do so, we used a series of cross-validated linear mixed effect models, where we examined the degree to which visual, linguistic, and multimodal similarity metrics (and their combination) derived from multimodal large language model embeddings could explain children's error patterns. Specifically, we modeled the proportion of time that children chose each distractor for a target word as a function of the age (in years) of the children participating, and (1) the similarity of the target word to each distractor word (linguistic embeddings), (2) the similarity of the target image to each distractor image (visual embeddings), and (3) the similarity of the target word to each distractor image (multimodal embeddings), and (4) a model with all predictors combined; we additionally added the total number of errors per item as a covariate, with random intercepts for each target word. We iteratively sampled 80 percent of the dataset 50 times, and then evaluated the conditional R-squared for each model for each split; these values are plotted in Figure 4C. These results revealed that combining both the visual and linguistic embeddings—either together in one, large mixed-effect model, or via multimodal embeddings—led to increased explained variance in children's error patterns. These results thus suggest that changes in children's error patterns across age are not solely due to changes in children's ability to reject the distractor images that are visually similar to the target concept. Rather, children's errors are driven by a combination of both visual and linguistic similarity.

Discussion

How precise is children's visual concept knowledge, and how does this change across development? Here, we collect and analyze a large dataset of picture matching performance across development, finding evidence for a transition from coarse to finer-grained visual representations over early and middle childhood. Children became more accurate at identifying the referents of words over this entire age range, and their error patterns progressed from relatively random towards related distractors.

Broadly, these data support a theoretical view where these is substantial enrichment and change in existing representations for everyday visual concepts throughout childhood. For example, certain visual features may become more or less salient in children's visual concepts as children understand their functional roles (e.g., camels have humps to store water) or the degree to which they help delineate a category boundary (e.g., between whales and whale sharks). On this account, even school-aged children's visual representations may undergo substantial change as they learn more about the world around them.

This protraction of the timeline for visual concept learning into middle childhood substantially broadens the scope of potential learning mechanisms beyond associative label-object matching. For example, children's learning environments extend beyond the home and into structured educational contexts; children's learning partners include their peers, teachers, and siblings (who may be more or less reliable), and children's individualized experiences, interests, and hobbies may influence which words they have detailed visual representations for. As children begin to learn why animals and objects are classified the way they are, this semantic learning likely influences the visual features that are prioritized in children's representations. To make matters even more complicated, children also likely learn about visual features from generic utterances (e.g., "Tigers have stripes") (Gelman, Ware, & Kleinberg, 2010) in verbal conversations where visual referents are nowhere to be found. Thus in order for our models and theories of visual concept learning to account for this full

334 developmental trajectory, we need to think beyond labelled (or even captioned) photos or

335 videos of referents.

336   Indeed, we suspect that visual concept learning extends into adulthood, and that many

337 adults have coarse visual representations for many different words (and indeed, we culled

338 some items during pilot testing because adults could not discriminate them!). Consider that

339 while many adults in Western contexts experience the referents of some visual concepts

340 relatively frequently e.g., trees, computers, cups, cars – other words refer to visual concepts

341 that different individuals may have varying amounts of interest in and frequency in

342 interacting with – like telescopes, or antelopes. Visual concept learning is likely influenced by

343 both individuals pre-occupations and very intense interests, be they professional or not. And

344 indeed decades of work has established that birding experts, car aficionados, (Tanaka &

345 Gauthier, 1997) and graphic artists (Perdreau & Cavanagh, 2013) have both qualitatively

346 and quantitatively different kinds of visual representations for the visual concepts that they

347 engage with.

348   There are several limitations to the current work that future work could address.

349 While we include data from a diverse group of children over a wide developmental age range,

350 at present our conclusions are drawn primarily from around one hundred experimental items

351 and distractors; further work that expands the range and diversity of the visual concepts –

352 and that expands to populations outside of the continental U.S.(Henrich, Heine, &

353 Norenzayan, 2010) will be necessary to understand the generalizability of these findings. In

354 addition, the present data are large but cross-sectional, and thus cannot provide evidence for

355 changes in the precision of representations within individual minds. Dense data collected

356 from children over longer ranges of developmental time could confirm the hypotheses and

357 theories raised by these analyses. Nonetheless, the present work highlights the promise of

358 large-scale, online games for collecting large datasets that can be used to examine the

359 consistency and variability in visual concept representations across childhood.

360   Overall, these findings suggest that children's visual concepts gradually become more

361 precise across childhood, and broaden our view on the timeline and mechanisms for visual

362 concept learning. We hope that future work will build on the tools and ideas developed here

363 to understand the visual mind in both children and adults.

³⁶⁴ **Acknowledgments**

³⁶⁹ **Appendix**

³⁷⁰ **References**

³⁷¹ Gelman, S. A., Ware, E. A., & Kleinberg, F. (2010). Effects of generic language on category

³⁷²   content and structure. *Cognitive Psychology*, *61*(3), 273–301.

³⁷³ Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Most people are not WEIRD. *Nature*,

³⁷⁴   *466*(7302), 29–29.

³⁷⁵ Perdreau, F., & Cavanagh, P. (2013). Is artists' perception more veridical? *Frontiers in*

³⁷⁶   *Neuroscience*, *7*, 6.

³⁷⁷ Tanaka, J., & Gauthier, I. (1997). Expertise in object and face recognition. *Psychology of*

³⁷⁸   *Learning and Motivation*, *36*, 85–126.

Table 1

Coefficients of a linear regression assessing changes in the proportion of related distractors chosen over development. Age and number of trials were standardized prior to analysis.

Predictor	b	SE	t	p
Intercept	0.62	0.01	123.63	< .001
Age (scaled)	0.04	0.01	7.45	< .001
Number of errors (scaled)	-0.05	0.01	-8.23	< .001
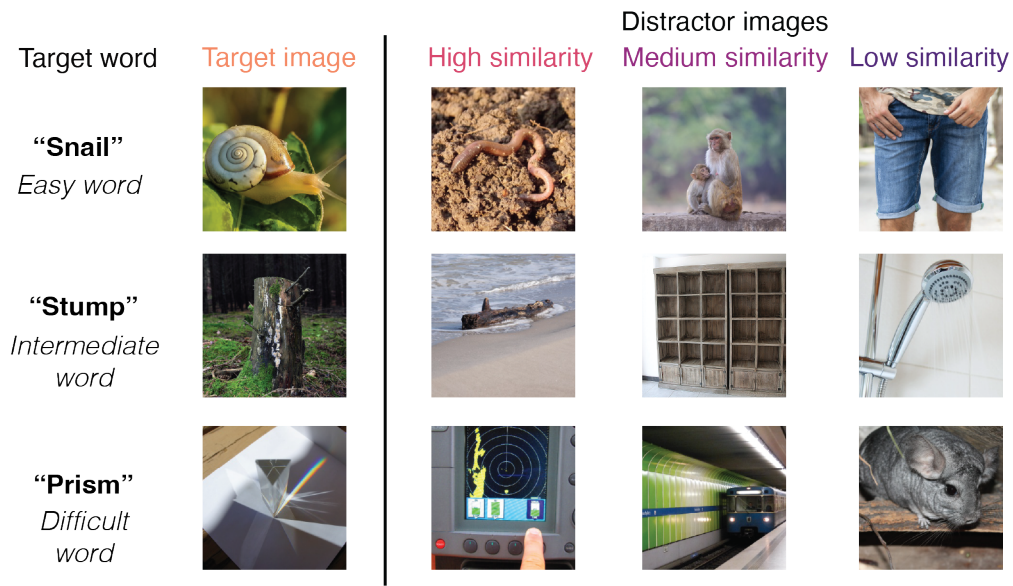
Figure 1. Example trials that depicted easy, intermediate, and difficult words, as operationalized via estimated age-of-acquisition (Kuperman et al., 2012); each target word image is taken from shown alongside the three distractors that varied in similarity to the target word; distractors were chosen via the similarity of the target and distractor words in the language encoder of CLIP (Radford et al., 2021).
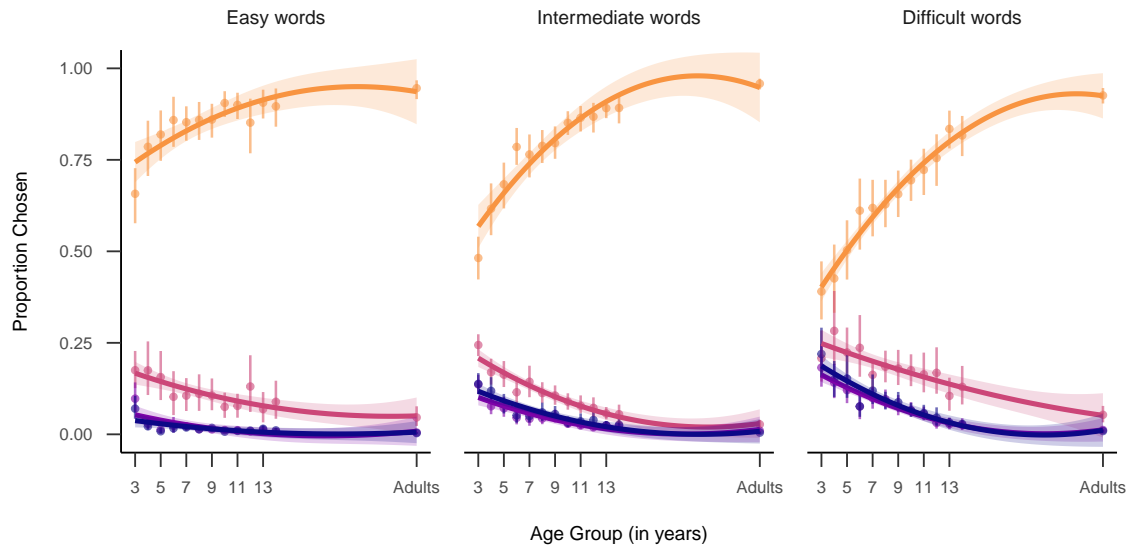
Figure 2. Visual vocabulary task performance as a function of the age of the child completing the task, plotted separately for relatively easy, harder, or difficult words; words are binned into terciles based on the estimated AoA from Kuperman et al., 2012. Lines refer to the proportion of words that children chose the target (red), high-similarity (green), medium similarity (turqouise), or low similarity (purple) distractor at each age; error bars represent boostrapped confidence intervals.
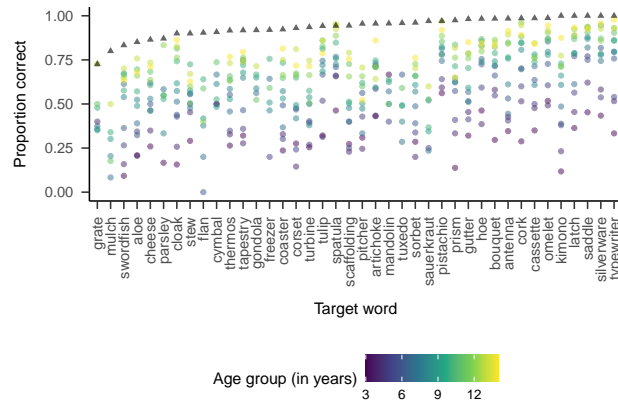
Figure 3. Visualization of the items that showed the greatest changes in distractor choice across age groups. Colors represent the age of the child participating in the task; triangles represent data from adults.
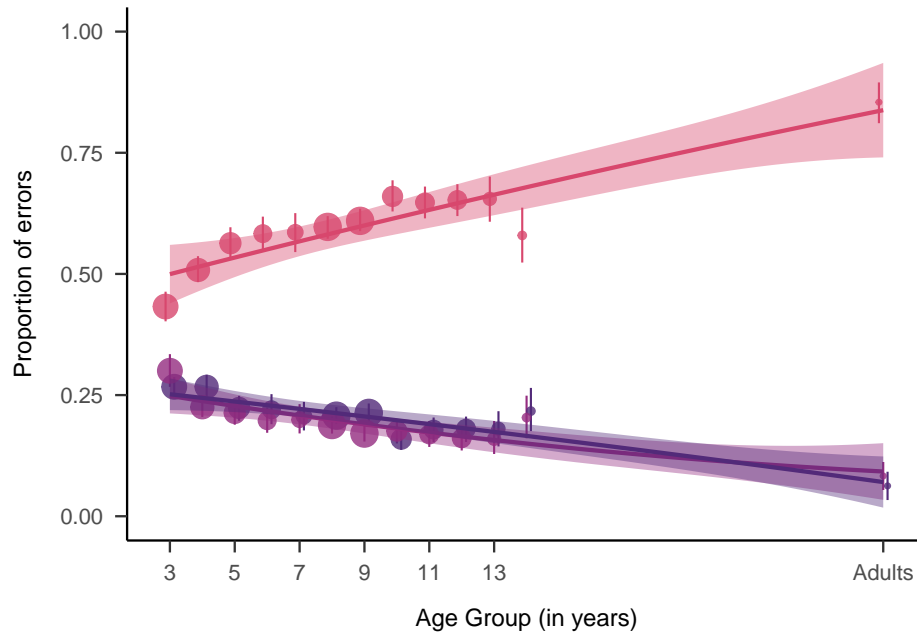

Figure 4. Changes in the proportion of errors chosen as a function of childrens age, where green lines reflected higher similarity distractors. Dot size represents the number of errors made by children in each age group. Error bars represent 95 percent bootrstrapped confidence intervals.
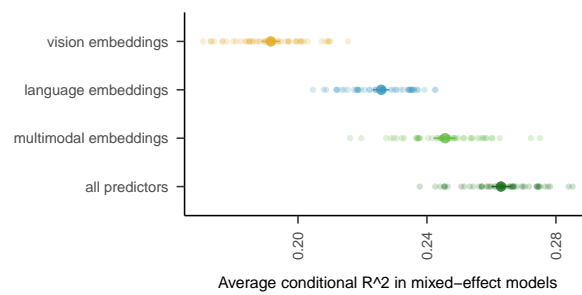
Figure 5. Average explained variance in children's error patterns (conditional R-squared in linear mixed effect models) by linguistic, visual, multimodal, or combined predictors in cross-validated mixed efffect models. Error bars represent bootstrapped 95 percent confidence intervals across 50 iterations.