Constructing Meaning: The Influence of Referent Experience on Infants' Word Knowledge

By

Haley Weaver

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Psychology)

at the

UNIVERSITY OF WISCONSIN–MADISON

2025

Date of final oral examination: 07/08/2025

The dissertation is approved by the following members of the Final Oral Committee:
Jenny R. Saffran, Rubinstein Professor, Psychology
Martha Wagner Alibali, Professor, Psychology
Karen B Schloss, Associate Professor, Psychology
Haley A. Vlach, Professor, Educational Psychology

Table of Contents

Acknowledgements

An acknowledgement section in a dissertation is too brief to adequately thank the wonderful community that made an impact on my graduate career over the past 6 years. Yet, I hope to impart my sincerest gratitude to some of the extraordinary forces who helped me achieve this accomplishment.

First, I want to thank the families who participated in this research, without whom, none of this work would be possible. You took an hour out of your week to sit in a silly little room with your child and listen to words repeatedly, just for the sake of scientific discovery. Your dedication to research is very appreciated.

I am also immensely grateful for the mentorship of my advisor, Jenny Saffran. Not only did you provide unwavering support and feedback along every step of my graduate career, but you also cared deeply about my well-being and personal growth. I cannot express how enjoyable it has been to discuss the minutiae of experimental design with such a careful and considerate scientist. I am so fortunate to have had the opportunity to collaborate with you for the past 6 years.

I also want to extend my sincerest gratitude for the supportive community of friends (who are also, fortunately, my colleagues) at the Infant Learning Lab and UW–Madison more broadly. Thanks to Aja Altenhof, Desia Bacon, Elise Breitfeld, Jess Mankewitz, Ron Pomper, Chenelle Walker, and Martin Zettersten for creating a tremendously stimulating scientific environment, but also one filled with kindness and encouragement. Thank you to my fellow graduate cohort friends with whom I've weathered a global pandemic, and so much more. To the Odyssey Junior family, you allowed me to truly see development in practice and provided joy every week. You opened my eyes to a mission that I hope to continue no matter where my path takes me. I appreciate every conversation and laugh we've shared, which have made this journey so enjoyable.

Lastly, I would like to thank my family for their support throughout my academic journey. You nurtured my inquisitiveness and always believed in my goals (even if you were not quite certain what my job prospects would ultimately be). A special thanks to my nieces and nephew, watching you grow and learn sparked my passion for language and development. Finally, thank you to Tom, who provided steadfast love, shared countless laughs, and spent a considerable amount of time editing my work. I will be forever grateful to Madison for bringing us together.

Abstract

Variability is ubiquitous in early language development. Infants' vocabularies vary in size and structure, and the referents present during individual word learning moments vary in number and kind. While variability in vocabulary has been studied extensively over the past two decades, variability in infants' individual word meanings has not been well examined. The current studies focus on one aspect of variability–differences in referent experience–to investigate individual differences in early word meanings. Study 1 characterized the nature of infants' referent experiences by developing a referent database using a novel method; caregivers provided and normed images that reflected their child's experiences. The database included 964 images from 40 different word categories and was normed on name agreement, image experience, and pairwise similarity. Across words and infants, referent experience varied substantially. Study 2 investigated whether variation in referent experiences had consequences for infants' word recognition. 9- to 11-month-olds and 15- to 17-month-olds were tested on their comprehension of food words–an ideal test case due to the high variability of food forms that infants encounter (e.g., whole, mashed, cut). Infants' word recognition was measured in a looking-while-listening paradigm with different forms of food (e.g., applesauce, apple slices). Caregivers also reported how often their child experienced each form of food in the past week. The older infants recognized words significantly above chance and were more accurate for referents that were reported to be very frequently experienced. The results suggest that infants associate known words more strongly with the referents that they frequently encounter. Taken together, the results demonstrate remarkable variability in early referent experience and provide preliminary evidence for individual differences in early word meanings.

Keywords: referents, familiar word learning, infancy, individual differences

Constructing Meaning: The Influence of Referent Experience on Infants' Word Knowledge

Infants learn words at a remarkable speed. Even before their first birthday, infants comprehend some highly frequent nouns, such as those referring to body parts, foods, and caregivers (Bergelson & Swingley, 2012, 2013, 2015; Campbell & Hall, 2022; Tincoff & Jusczyk, 1999). These comprehension abilities improve exponentially across the second year of life (Bergelson, 2020; Fernald et al., 1998) as toddlers begin to restructure their burgeoning vocabularies to represent similarities across word meanings in rich, interconnected networks (Arias-Trejo & Plunkett, 2010; Borovsky, 2022; Borovsky et al., 2016b; Mani & Plunkett, 2011; Peters & Borovsky, 2019). Not only do toddlers' word meanings include semantic relationships, but they also extend to both typical and atypical category members, suggesting robust adult-like representations (Meints et al., 1999; Southgate & Meints, 2000; Weaver et al., 2024).

Historically, studies of word comprehension have employed looking-while-listening (LWL) paradigms to probe infants' ability to link a *single* prototypical referent with a label in the presence of a single prototypical distractor, at the group level. This reliance on label-referent mappings has recently been coined the "map trap". Under this framework, infants' primary word-learning challenge is to link a word with a "core" referent to be determined by adult norms or researchers' intuitions (Wojcik et al., 2022). However, infants must include entire categories of referents that may vary in typicality and/or form to learn the meaning of a word. For instance, infants may encounter yellow unpeeled bananas, round beige banana slices, or beige mashed bananas when they hear the word "banana". In addition to contending with exemplar variability, infants must also be able to recognize the referents of words in the presence of a variety of other (often related) objects. That is, bananas co-occur with high-chairs, spoons, bowls, bibs, and even other foods (Clerkin et al., 2017; Clerkin & Smith, 2022). These examples highlight the complexity and variability in real-world word learning that traditional LWL paradigms fail to adequately capture.

The “map trap” framework also presupposes that the single referent is representative of infants’ underlying word meanings. That is, researchers designing experiments have to select images that they expect infants to link with labels. However, there is limited evidence that the images used in experiments are the actual referents that infants encounter and associate with words. For example, a recent study demonstrated that infants failed to recognize highly frequent, early learned abstract words (e.g., uh-oh) when using researcher-selected images (Casey et al., 2023). The researchers examined a set of naturalistic home recordings that included their target words and found that the real-world referents that infants experienced were vastly different than those tested in the experimental paradigm, suggesting that adult intuitions about appropriate referent images may not align with infants’ representations. Therefore, it is possible that prior studies of early word comprehension are not adequately, or potentially even accurately, testing infants’ word representation as a result of the specific stimuli used (i.e., referents selected based on adults’ word meanings).

While infants’ word meanings have been largely reduced to a small set of prototypical label-referent mappings, mature (i.e., adult) word representations include a vast array of referents and related semantic features (McNamara & Sternberg, 1983; McRae et al., 1997; Wulff et al., 2022). That is, adults’ word meanings are not ubiquitous, stationary targets that infants must acquire. Instead, word meanings are idiosyncratic and multi-faceted, reflecting in the moment sentential context as well as a wealth of lived experience with language and the world specific to the listener (Ramsey, 2022; Rodd, 2020; Rodd et al., 2016; Thompson et al., 2020; Wang & Bi, 2021; Wiley et al., 2018). Prior domain knowledge and recent experience, in particular, seem to influence adults' interpretations of specific word senses or meanings. For example, adult rowers were more likely to interpret ambiguous sentences that included words such as "square" in line with the rowing sense (i.e., an oar position) rather than the dominant sense (i.e., the shape; Rodd et al., 2016). Similarly, adults who played baseball did not incorporate prior sentential context into their interpretation of ambiguous sentences with baseball-relevant terms (i.e., "bat"). Instead, they were strongly biased to interpret "bat" in the baseball sense (Wiley et al., 2018). These examples demonstrate that adults' word representations are strongly influenced by their idiosyncratic experiences with the world, suggesting that the target of lexical development is not a single, shared set of conventional meanings.

Research on infants' vocabulary development provides hints that individual differences may also influence early word representations. Prior category knowledge appears to influence infants' understanding of words. Toddlers who have more shape-based categories of words in their vocabulary (e.g., cup, ball, shoe, etc.) are faster to identify an atypically colored, same-shape referent than toddlers with fewer shape-based words (Perry & Saffran, 2017). This finding suggests that some infants may prioritize shape in their word representations more strongly than others, depending on their vocabulary structure. Two-year-olds also recognize and/or learn words more readily from categories in which they know more words (Borovsky et al., 2016b, 2016a) and from categories in which they are highly interested (Ackermann et al., 2020). For example, a toddler who knows many food words is likely to be faster and more accurate at recognizing the referent of a food word compared to a word from another category. The observation that differences in word learning and recognition depend on prior category knowledge suggests that toddlers' word meanings are likely influenced by prior knowledge, just as adults' meanings are subject to domain expertise.

Prior work suggests that there may be individual differences in the word representations of young infants, and these may be sensitive to factors that influence adults' meanings, such as experience. These individual differences, however, remain largely unexplored due to two methodological limitations. First, LWL studies typically test a limited number of adult-relevant referents rather than including target referents that may be more appropriate given infants' real-world experiences. Second, language development researchers have limited empirical knowledge about the kinds of referents that young infants regularly experience. In the subsequent sections, we review the literature regarding the role of experience on word learning and how we, as a field, can leverage caregiver report methods to better measure experience relevant to word representations.

Referent Experience and Words

Word learning involves dynamic interactions between infants, language input, and the world. To learn a word's meaning, an infant must be able to recognize the real-world referent and associate it with a label. Researchers have puzzled over infants' ability to identify the referent for a label for decades since Quine (1960) first described the induction problem. One proposed solution to this problem is infants' ability to track statistical co-occurrences between words and referents across individually ambiguous instances (Smith & Yu, 2008). That is, infants can use occurrence statistics about how often a word is uttered in the presence of different objects to select a likely referent. Studies conducted in laboratory settings have demonstrated strong evidence that cross-situational word learning is a viable mechanism for understanding which words map to which referents (Crespo et al., 2024; Nikolaus & Fourtassi, 2021; Vlach & DeBrock, 2019).

Prior experiments have primarily interrogated cross-situational word learning in a simplified version of infants' actual word learning experience. Infants are exposed to a single exemplar for each of the referents during a cross-situational word learning task, barring one exception: Crespo and colleagues (2024) demonstrated that children (5- to 8-year-olds) can learn words in cross-situational paradigms that use variable exemplars. In the real world, however, even very young infants experience a variety of different exemplars when learning words. For instance, an infant may hear the word "dog" in the presence of a golden retriever in the backyard or in the presence of a big red dog when reading a book. Critically, infants likely experience these exemplars with different frequencies. Given infants' robust statistical learning abilities (Saffran & Kirkham, 2018), it is possible that infants are tracking the individual exemplars that co-occur with a given word in addition to the labels that co-occur. The variation in the frequency of different exemplars could, in principle, lead to different word meaning representations, both across infants and developmental time. We aim to both understand the breadth of variability in referent experiences as well as understand whether experience may shape word meanings.

While the specific referents that infants experience remain understudied, a growing body of literature highlights visual object experience as a critical component of word learning. For example, the earliest learned words are those whose visual referents are readily available during infants' mealtime experience (Clerkin et al., 2017) despite their labels being relatively infrequent (Clerkin & Smith, 2022). Object experience alone facilitates infants' word extension to new exemplars (Pomiechowska & Gliga, 2019) and improves retention for new word-referent mappings following a short delay (Kucker & Samuelson, 2012). Furthermore, visual experience with referents is highly variable across contexts (Custode & Tamis-LeMonda, 2020). The extent to which labels co-occur with their referents also varies; labels are heard in the presence of referents for certain words, but for other words, referents are often absent (Clerkin & Smith, 2022; Wojcik et al., 2024). Together, these findings suggest that word learning is facilitated by prior exposure to referents despite the highly variable and word-specific nature of referent experience.

Descriptive studies of infants' word learning environments have been critical in understanding how infants may learn words, but they have not yet uncovered how these variable experiences inform infants' representations of words. Specifically, naturalistic referent experience has largely been ignored in experimental tasks designed to study infants' early word meanings. Only a few prior studies have considered the role of experience in infants' word recognition. Garrison and colleagues (2020) tested word recognition in the presence of exemplars from the infants' own home (i.e., a photograph provided by their caregiver) and a prototypical new exemplar of a familiar word (i.e., never seen by the participant). Older infants were equally accurate at recognizing familiar and new referents for familiar words. However, younger infants only recognized familiar referents significantly above chance. This suggests that referent experience is especially important when examining the word meanings early in development. Lending further credence to this hypothesis, Casey and colleagues (2023) found no evidence of comprehension of abstract words reported to be produced by most 16-month-olds (e.g., all done, uh-oh, and bye-bye) when using researcher-selected referents in an LWL task. After examining naturalistic videos of infants' experiences, researchers determined that the visually-available referents were different from those that the researchers had selected. In particular, referents for abstract words were more variable and referred to situational contexts rather than a single concrete event (e.g., "shh" was said most often in the presence of someone sleeping, screaming, or crying). These findings underscore the importance of understanding early referent experience and using more ecologically-valid stimuli to investigate infants' representations of words.

While Casey and colleagues (2023) used naturalistic home video recordings to examine experience-relevant referents, this method is not a tractable solution for word recognition studies using LWL. Stimuli mined from naturalistic videos are severely limited by the quality and quantity of the video data available. Critically, studies that record videos of the home word learning environment only sample very specific, and potentially unrepresentative, moments of infants' lives (Bergelson et al., 2019; Fausey, 2022). As an alternative, the next section proposes caregivers' reports as a critical source of word-referent experience.

Caregiver Report

Caregiver reports of word comprehension are routinely used in language development research (e.g., Bennetts et al., 2016; Bricker et al., 1999; Fenson et al., 2007; Rescorla, 1989). Caregivers are uniquely positioned to provide insight into their child's language abilities because they can observe comprehension during naturalistic environments across many different contexts. By comparison, clinical and research interactions are limited in that they typically take place in unfamiliar settings with unfamiliar interlocutors influencing children's performance (Miller et al., 2017).

One prevalent caregiver report tool used in research settings is the MacArthur-Bates Communicative Development Inventories (MB-CDI; Fenson et al., 1994, 2007). This measure was developed as a normative assessment of vocabulary development across the first two years of life (Fenson et al., 1994, 2007). Although originally developed in American English, this tool has been adapted to many other dialects and languages (e.g., Jago et al., 2023; Maital et al., 2000). The MB-CDI has been foundational in understanding the size and structure of infants' growing vocabularies (Mayor & Plunkett, 2011, 2014; Perry & Samuelson, 2011; Peters & Borovsky, 2019; Samuelson & Smith, 1999) as well as providing key insights regarding variability in language development (Marchman & Fernald, 2008a; Peter et al., 2019; Slone et al., 2019; Yu et al., 2019). Due to its widespread use, researchers have also developed a large-scale repository of MB-CDI data, WordBank, which provides evidence-based trajectories of the words that infants comprehend and produce at different ages (Frank et al., 2017).

Despite broad use in child language research, the validity of caregivers' reports has been called into question, particularly for the words that caregivers report infants to understand. Caregivers' reports of word knowledge accurately reflect infant behavior on some experimental tasks (Ring et al., 2000; Styles & Plunkett, 2009; Syrnyk & Meints, 2017), but they often underestimate word comprehension compared to eye-gaze behavior during LWL tasks (Bergelson & Swingley, 2012; Houston-Price et al., 2007; Lorang et al., 2023; Valleau et al., 2018; Venker et al., 2016). One proposed explanation for differences in infant performance on word comprehension and caregiver report is that the tasks have fundamentally different designs (Weaver & Saffran, under review). LWL tasks assess a simplified form of word comprehension by which infants must identify a specific referent as the target of a label in the presence of one (often unrelated) distractor. By contrast, caregiver report methods draw on caregivers' accumulated knowledge of their child's experiences with words and referents, asking them to synthesize this rich data into a single response. A caregiver may consider whether their infant understands a prototypical "adult" meaning of the word "apple" (i.e., a shiny red apple). Alternatively, a caregiver could consider whether their infant recognizes an object when it appears in various forms (e.g., sliced, sauced, juiced, and green apples) as the meaning of the word "apple". They may even be considering a specific interaction as evidence that their child understands a word (e.g., their infant points to an apple when queried). Each of these considerations could result in dramatically different reports of the words that are understood by a given infant, especially when compared to what researchers consider to be word understanding in an LWL experiment (i.e., looking more at a labeled object).

Although caregivers' considerations of various word learning situations may lead to more variation in reporting their infants' word knowledge, it may also improve their understanding of other aspects of language development. Caregivers endorse research-based word learning mechanisms as ways infants learn words, despite the broader adult population's belief that some of these mechanisms are too challenging for young infants. For example, parents reported that preschoolers could use cross-situational word learning to learn the meaning of a newly presented word, whereas undergraduates and other adults did not (Knabe et al., 2023). Parents also adapt their input to their child's language development abilities. For instance, Kosie and Lew-Williams (2024) demonstrated that caregivers used more gestures and emotional cues in their infant directed communication with infants who had small vocabularies, potentially scaffolding word learning for those with weaker vocabulary skills. Leung and colleagues (2021) similarly found that parents provided more information in their referential expressions for words that their child did not know (e.g., "the spotted yellow cat-like animal" to refer to a leopard). These findings demonstrate clear evidence that caregivers have a wealth of knowledge about their child's language development. Researchers, however, may need to develop creative methods to tap caregivers' insights.

One potential method for probing caregivers' dense knowledge of their child's language is using specific questions that are grounded in observable behavior. For example, caregivers appear to evaluate word learning mechanisms when presented with specific vignettes (Knabe et al., 2023). They are also more accurate at reporting productive vocabulary than receptive vocabulary, likely because identifying an instance of spoken language is more obvious than an instance of comprehension. Word-referent experiences may be an additional area that caregivers could easily observe during everyday interactions with their infants. It is, therefore, possible that caregivers would be particularly good at reporting the kinds of referents their infant has experienced.

Research Aims

We aim to address two understudied aspects of infants' early word meanings: (1) to determine the range of referents that infants regularly experience for a set of common nouns and (2) to examine how infants' individual and developmental experience with different referents shapes early word meanings. Study 1 addressed Aim 1 by creating a database of referent images provided and normed by caregivers. In Study 1a, we collected images by asking caregivers to upload photographs of the objects their child regularly experiences for a set of frequently studied nouns. Then, in Study 1b, a separate set of caregivers rated the images from Study 1a on several dimensions that could be potentially relevant for LWL studies (e.g., name agreement, experience, and similarity). Study 2 addressed Aim 2 by measuring word recognition in 9- to 11-month-olds and 15- to 17-month-olds, focusing on food words in the presence of referents that are appropriate for two different age groups (e.g., shredded cheese vs. sliced cheese). We used food words because we expected substantial variability in the forms that infants experience for these referents, providing an opportunity to probe the effects of real-world experience on lexical representations (including potential individual differences). Caregivers also reported their infant's experience with different words and referents to provide a measure of individual differences. Together, these experiments shed light on whether and how referent experience varies across infants and development.

Preliminary Study A

In previous work, we examined the breadth of early word meanings for a set of familiar animal nouns. Specifically, we investigated whether 14- to 18-month-olds extended their word meanings to include both typical (e.g., golden retriever) and atypical (e.g., basset hound) category members. We also measured whether caregivers' reports of their child's experience with each category member explained individual differences in word recognition. We tested infants' word recognition using an LWL task including a variety of different category members (6 referents per word) for a small set of animal words (*dog*, *cat*, *bird*, and *fish*). To measure each infant's prior experience, we asked caregivers to rate whether each image was typical of their infant's experiences with the broader category (e.g., *"How typical is this image of a beagle of the dogs your child experiences?*"). The results indicated that infants' word recognition was remarkably robust across both exemplar typicality and individual differences in experience. 14- to 18-month-olds extended words to highly typical and atypical category members, demonstrating that word meanings refer to broad categories of referents early on in development (Weaver et al., 2024).

The previous results underscore the importance of moving beyond simple one-to-one mappings between a word and a referent when examining early word meanings. By including a wide variety of referents, we interrogated infants' underlying meaning representations, finding that even 14-month-olds include exemplars that they have likely never experienced before (e.g., kookaburra for "bird") in their word meanings. Although this study found no evidence for experience-based individual differences, several open questions remain. For example, we cannot rule out the possibility that younger infants' word meanings are narrower and more idiosyncratic due to relatively less experience with words and their referents. The present findings were also limited to animal words. Animal referents may be more perceptually similar than referents from other word categories, ultimately facilitating infants' word extension. Thus, the current studies aim to capture referent experience across a broad development age range as well as to investigate meanings in younger infants for a new category of words.

Preliminary Study B

In a separate line of research, we aimed to examine the validity of two gold-standard measures of word knowledge in infant research: caregiver reports of vocabulary and LWL tasks. Prior studies examining the convergent validity between these two measurements have been largely inconclusive. Some researchers demonstrate that caregivers underestimate their infants' understanding of individual words (Bergelson & Swingley, 2012; Houston-Price et al., 2007; Lorang et al., 2023; Venker et al., 2016), while others demonstrate that caregivers are accurate reporters of infants' word comprehension when compared to LWL tasks (Ring & Fenson, 2000; Styles & Plunkett, 2009; Syrnyk & Meints, 2017). To delve deeper into this set of questions, we investigated infants' comprehension of eight familiar words in an LWL paradigm and compared the speed and accuracy of infants' target looking to their caregiver's report of knowledge (i.e., said, understood, or unknown) for the same words (Weaver & Saffran, under review).

In addition to examining convergent validity between caregivers' reports and LWL, we were interested in understanding potential reasons (i.e., moderators) to explain why these two measurements may not be aligned when considering infants' word knowledge. We focused on moderators that would ameliorate design differences between the two measures. Specifically, we created a more robust and ecologically-valid LWL paradigm that tested multiple target images in the context of multiple distractor referents per word. We reasoned that caregivers are likely estimating their child's word knowledge across multiple contexts and in the presence of many referent objects, and thus, we wanted the LWL task to more closely resemble this natural variability. We also collected caregiver reports of their confidence in their estimates of word knowledge because caregivers may be attuned to their own reliability.

Caregivers' reports and infants' LWL behavior were more aligned when we considered relevant task features. Specifically, caregiver reports of infants' knowledge of *both* the target and distractor word significantly predicted infants' gaze behavior (Weaver & Saffran, under review). Put another way, infants' word recognition was superior for trials that included a target word or a distractor word that their caregiver reported as known. Similarly, caregivers' confidence in their ratings was a significant predictor of alignment. When caregivers were more confident, reports of word knowledge were more likely to be aligned with infants' word recognition. By considering the task features for both measures, we may have more closely equated behavior in each task. That is, word knowledge is aligned across methods when we account for the context of word recognition behavior and caregivers' certainty in reporting. These findings suggest that caregivers are a valuable resource to understand infants' word knowledge. They also demonstrate that we can measure word knowledge with greater validity by considering the design of the tasks. Ultimately, these results provide additional support for using caregiver report methods in future experiments regarding word knowledge, as their reports align with more direct measures of behavior under the right circumstances.

Study 1

Visual stimuli are a critical component of interrogating early receptive vocabulary. Experimental tasks such as the visual world paradigm (Huettig et al., 2011), the intermodal preferential looking paradigm (Golinkoff et al., 2013; Hirsh-Pasek et al., 1987), and the looking-while-listening paradigm (Fernald et al., 2008) all examine associations between speech and images. During these tasks, researchers measure the speed and accuracy of participants' fixations to images as a speech stream unfolds. Critically, at least one of the images displayed is related to the speech, while the other(s) are not. Thus, participants' visual attention should be directed to the associated image by the speech stream, as long as they recognize a given image as a referent of a specific word.

Researchers must consider several factors when selecting visual stimuli to probe language comprehension. The images could be color or greyscale photographs, line drawings, or illustrations; this choice of design features influences young children's ability to transfer knowledge between 2D and 3D representations (Ganea et al., 2009). Beyond the style of the images, researchers must also decide on the specific example of the object to be included in the images. For example, an image of chocolate, vanilla, red velvet, or even carrot cake could be included in an experiment testing comprehension of the word "cake". A given experimental task could include a multitude of different images, so how do researchers decide which visual stimuli to use?

One potential solution is to select images from clinical assessments of receptive vocabulary (e.g., Peabody Picture Vocabulary Test; Dunn & Dunn, 2007), which include images (and words) that are extensively normed to ensure that typically-developing participants identify and link the depicted referent with its intended label (Adlington et al., 2009; Dunn & Dunn, 2007). However, stimuli from clinical assessments are not commonly used in research settings as they include particular kinds of words (e.g., words that classify different language abilities). In tasks with adults, researchers have drawn images from normed sets of visual stimuli (e.g., Hintz et al., 2017; Huettig & Altmann, 2005; Zarcone & Demberg, 2021). Infancy researchers, however, often select their images based on their intuitions about word meanings.

Selecting images that are relevant to infants' word meanings is a challenging feat. Infants' visual experiences are diverse and potentially quite different from adults' visual experiences (Casey et al., 2023; Clerkin et al., 2017; Clerkin & Smith, 2022). Indeed, scenes that adults intuit to be highly cluttered are simplified by infants' object manipulations and attention (Slone et al., 2019, 2023). Infant-specific visual statistics and experiences, however, are argued to be precisely the kinds of experiences that facilitate word learning (Chen et al., 2021; Clerkin et al., 2017; Clerkin & Smith, 2022; Slone et al., 2019; Suanda et al., 2017, 2019; Yu et al., 2008, 2019). While there is strong evidence suggesting that the visual ecology of early word learning differs from adults' intuitions, researchers do not typically consider how this variability may influence infants' recognition of words in the presence of different kinds of stimuli (c.f. Casey et al., 2023; Garrison et al., 2020). Indeed, there are very few studies that have tested referents that are taken from scenes directly experienced by infants (Garrison et al., 2020), potentially due to the limited availability of these kinds of images in existing databases.

While many image databases are aimed at and normed by adults (Brodeur et al., 2010, 2014; Hebart et al., 2019; Krautz & Keuleers, 2022; Migo et al., 2013), there are fewer developed directly for use with young children. Several existing adult stimulus sets have been normed with children (i.e., ages three- to eight-years-old; Barry et al., 1997; Berman et al., 1989; Cannard et al., 2005; D'Amico et al., 2001; Masterson et al., 2008; Pompéia et al., 2001;

Sommerfeld et al., 2022). However, there have been no databases of familiar objects developed specifically for use with children, or designed to investigate infant development (i.e., children younger than three years). The only database that is frequently used by developmental scientists is the Novel Noun and Unusual Name Database (Horst & Hout, 2016). This database includes *novel* words and images rather than images of objects that are typical of infants' real-world language experience. Thus, there is a critical need for images and norms of familiar objects that are adequate for use in experiments with infants and toddlers. The present study aims to address this gap by developing a new set of image stimuli that are aimed at and normed for developmental science.

In developing the database, we had several important considerations. First, to ensure that the database is broadly relevant to many different developmental questions, the images and norms must be informed by prior research with infants. For this reason, we focused on referents whose labels are frequently studied by developmental scientists. Secondly, the images should reflect infants' lived experiences with referents. Finally, the norms should measure factors known to influence lexical processing. In the next sections, we review each of these considerations in turn.

Early-learned Words

The MacArthur-Bates Communicative Development Inventories (MB-CDI; Fenson et al., 2007) are a tool frequently used to assess infants' language development. The MB-CDI is a standardized vocabulary checklist including over 600 words and is intended for 8- to 30-month-olds. Importantly, there are several versions of the checklist that include different words depending on the age range for which the form is intended. For example, the Words and Gestures Short Form is normed for 8- to 18-month-olds and only includes 100 words across categories such as food, toys, furniture, games and routines, and people. Due to its widespread use in the field of child development, researchers can investigate acquisition trajectories for specific words averaged across thousands of participants (Frank et al., 2017). This dataset allows researchers to select age-appropriate words depending on the question of interest. As such, the words on the MB-CDI are frequently tested in experimental settings (e.g., Bergelson & Swingley, 2012b, 2013, 2015; Casey et al., 2023; Garrison et al., 2020; Weaver et al., 2024).

For our current purposes, the broad use of vocabulary checklists suggests that the words included on the MB-CDI are a worthwhile starting point for developing a database of images. Since this is the first database of its kind, we began by collecting images for a set of concrete nouns drawn from the Words and Gestures Short form. We reasoned that the selected words would be imageable (i.e., easily captured by a photograph) and would be of great interest to language researchers (Casey et al., 2023; Wojcik et al., 2022).

Experience-relevant Images

Most prior databases have been curated using researchers' intuitions and subsequently normed by the intended population. However, infants' visual experience with objects is vastly different from researchers' expectations. Thus, selecting images prior to norming likely will not yield a set of images that are relevant to infants' lived experiences. A possible alternative is to draw on infant methods (i.e., caregiver report) to amass images. While this method is highly atypical of stimuli databases, Garrison and colleagues (2020) asked caregivers to provide images of objects in their home, which were then used as stimuli in an LWL study, demonstrating the potential feasibility of this method. Ultimately, this is the method we chose to use for the first phase of database development (Study 1a).

Factors influencing comprehension

There are a multitude of important factors that may influence infants' in-the-moment word comprehension. As this is an initial examination of experience-relevant images, we focused on three factors that are likely to be relevant for an image database intended for use in different experimental designs: word form, familiarity, and similarity.

One commonly studied factor is the label, or word form, that infants associate with known objects. Word learning involves not only learning the sounds of words, but also learning the relationship between words and objects (Samuelson & McMurray, 2017). Indeed, infants hear a variety of word forms (i.e., dog vs. doggy) in their language input (Moore & Bergelson, 2024). They are also sensitive to different word forms and demonstrate slowed word recognition when words are mispronounced or less frequent (Potter & Lew-Williams, 2023; Swingley & Aslin, 2000). These findings highlight the importance of collecting and including naming information in a new database.

Prior experience, or familiarity, with a word's referent may also play a role in word comprehension. Given the variability in infants' visual experiences, the different kinds of objects that infants see in the presence of early words may influence their word representations. Indeed, 12-month-olds' word comprehension is boosted in the presence of photographs of objects from their own homes (i.e., their own shoe; Garrison et al., 2020). Older infants, however, robustly recognize words regardless of the familiarity or typicality of the referents (Garrison et al., 2020; Weaver et al., 2024). These results suggest that prior experience may have different impacts on word recognition at different ages. Experience is, therefore, a highly relevant variable to consider when creating new stimuli.

Finally, similarity between target and distractor images influences lexical processing. Infants are slower to identify a target image after hearing a label when the target is perceptually similar to the distractor (Arias-Trejo & Plunkett, 2010; Ellis Weismer et al., 2016). For instance, infants are slower to recognize the word "banana" in the presence of a banana and a yellow sock. These slowed processing effects have been interpreted as evidence that infants encode perceptual information about a word's referent (Arias-Trejo & Plunkett, 2010). Because infants represent perceptual similarity, many studies of familiar word comprehension intentionally pair highly dissimilar target and distractors (i.e., referents from different categories) to reduce task demands (Bergelson & Swingley, 2012b, 2015; Kartushina & Mayor, 2019). Taken together, this work demonstrates the value in providing standardized similarity ratings for a set of image stimuli.

In the present study, we developed a database of naturalistic referent images for imageable nouns on the MB-CDI that are representative of infants' experiences and rated on characteristics that could influence infants' word comprehension. First, we collected photographs from caregivers of the kinds of referents their child commonly experiences for a set of concrete, imageable nouns from the MB-CDI Short Form Level I (Study 1a). A separate set of caregivers provided labels for the images from Study 1a and rated how often their infants experienced the images as noun referents to develop a set of norms for each of the images (Study 1b). In addition to the caregiver surveys, Study 1b employed large vision models to create a metric of pairwise similarity between the images. The resultant image database and norms will be shared with the research community via the Open Science Framework.

Study 1a

Method

Participants

The final image set included data from 33 caregivers with at least one child under the age of 2 years. On average, caregivers had infants that were 13.2 months old (SD=6.9; Range = [1.2, 24.23]). An additional 5 caregivers participated but were removed from the final sample due to insufficient data contribution (N=2), for failing to pass the attention checks (N=1), or for providing images from the internet (N=2). We recruited caregivers with infants under 2 years of age because vocabulary is particularly difficult to assess in younger age groups (Fernald et al., 2006; Marchman et al., 2023) and experience with referents is likely to be more idiosyncratic for younger infants than for older children (Campbell & Hall, 2022; Clerkin et al., 2017; Clerkin & Smith, 2022; Garrison et al., 2020). This age range also overlapped with the intended age range for the vocabulary checklist from which we drew the list of nouns (8- to 18-months). The sample was limited to participants who indicated that they had a child born in 2024 or 2023. Participants also had to reside in the United States and had to report English as one of their fluent languages to be included in the sample, as the nouns were drawn from a vocabulary checklist intended for use with monolingual American English–learning infants.

The sample size for this study was selected based on the number of images typically included in adult databases (i.e., 147 to 1,680 images; Adlington et al., 2009; Brodeur et al., 2014; Krautz & Keuleers, 2022). We stopped collecting data when every object had at least 10 images provided, as this would yield a minimum of 400 images. Ultimately, our sample size yielded 1043 unique images.

Procedure

Participants completed a two-phase online survey hosted on Prolific (www.prolific.com; Douglas et al., 2023; Peer et al., 2022). The first phase included an eligibility screening questionnaire that asked participants to provide their infant's age in months. If the caregiver reported that their child was 24 months old or younger, they were then provided with a link to complete the second phase of the survey. Sixty-five participants completed phase one of the survey, and 59/65 participants were deemed eligible in phase one.

In phase two of the survey, the 59 eligible participants were provided with a link to complete a survey. The survey provided participants with a list of 40 concrete imageable nouns (see full list of words in Appendix A). Participants were instructed to select at least 30 words from the list and provide an image for each. The list of words was selected from the MB-CDI Short Form Level I and included 3 additional food words tested in Study 2 (which appear on the MB-CDI Words and Gestures long form). Body-part and people words were excluded to reduce the likelihood of identifiable information being inadvertently included in the photographs. The included nouns span a wide range of categories, including foods, animals, and household items.

The survey instructions highlighted the importance of data quality. For instance, participants were asked to upload images that centered the object and that were illuminated well enough to easily identify the photographed object. The instructions also requested that the images depict a referent that the participant's child is likely to experience when they hear the word of interest (e.g., "*Think about what your baby actually sees. For example, even though you might see the car you drive, your baby might see a toy car more often. We want you to send us pictures of things that are representative of your child's experiences.*"). Similarly, caregivers were urged to consider all the contexts in which their child experiences the word (e.g., "*Please* *think about all the media (books, TV, iPad, etc.) and real-world experiences (toys, mealtimes, zoo, etc.) your child has with each of these words.")*. However, we asked participants to refrain from providing copyrighted images to minimize challenges when reproducing images for an open-source database.

Caregivers also provided demographic information. They reported their race, ethnicity, education level, household income, and their child's age. These questions also served as a quality check to ensure that the adults contributing data have children in the correct age range and that they are not computerized bots. Participants were required to enter their infant's date of birth (DOB) at the beginning of the survey and indicate their infant's age in months at the end of the survey. Participants were excluded if their infants' age fell outside of the predetermined range (0- to 24-months) or if the reported DOB did not align with the child's age.

Participants were given a week to upload the images and complete the survey. We included a protracted timeline for submission to increase the likelihood that participants would photograph objects that their child encountered throughout the week.

Results

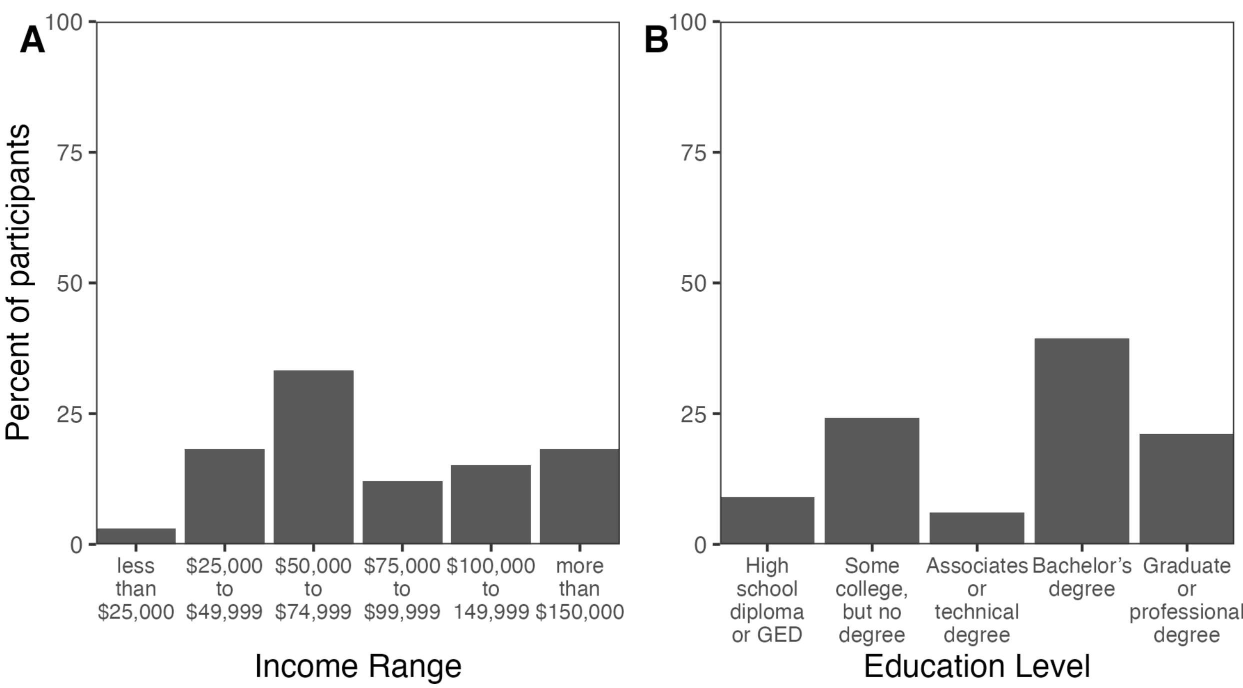
In Study 1a, we were interested in collecting a large set of images, which were normed in Study 1b. Thus, in the present study, we only report descriptive statistics for the images and demographic information about the participants.

In total, we collected 1,043 images across all participants. The number of images provided for each object word varied. On average, we received 26 images per object word (*SD*=6; Range= [10, 33]). Every participant (*N*=33) provided an image for the word "car", while only 10 participants provided images for the word "moon". The number of images provided for each word is included in Appendix B.

The demographics of the sample are important to consider as they provide context for the kind and range of images collected. Caregivers predominantly identified as White and not Hispanic (72%). Fifteen percent of the sample identified as Black, 3% identified as Asian, 3% identified with multiple racial categories, and 6% identified as White and Hispanic. In general, participants were middle class and educated. The most commonly reported annual household income range was $50,000 to $74,999 (Range= [less than $25,000, more than $150,000]), and most participants reported holding a bachelor's degree (N=13). However, our participants reported a range of annual incomes and education levels (Figure 1).

Figure 1

Demographic Information



Note. Panel A depicts the proportion of participants who reported each level of income. Panel B depicts the proportion of participants who reported each education level.

We also investigated whether the demographics reported in the current study are representative of the broader U.S. population by comparing our sample to the U.S. census reports. We gathered data regarding race and ethnicity as reported on the 2020 U.S. census (Table 1; U.S. Census Bureau, 2020). We also calculated the proportion of the families that reported earning each income range and the proportion of adults over the age of 25 who held each education level drawn from the 2023 American Community Survey (Table 1; U.S. Census Bureau, 2023, 2023). The census data reports educational attainment in binned age groups from 18 to 24 years and 25 years of age and older. Our participants ranged from 24 to 46 years of age, and so we calculated education levels using the data for adults over age 25, as it approximately matched the ages represented in our sample. In general, our sample had a higher proportion of participants who reported being White and Non-Hispanic than the national average (i.e., 72% vs 57.8%). Additionally, the current sample included far fewer Hispanic adults (6%) than the general US population (18.7%). However, our participants matched the range of educational attainment reported on the U.S. Census, with most adults in both samples reporting a bachelor's degree. With respect to income, our sample was less affluent on average than the U.S. population. The modal income range for the sample was $50,000 to $75,000, while the national modal income range was over $150,000. It is important to note that we did not collect information regarding the size of households for the current sample. Thus, we were unable to compare median incomes relative to the average household size in our sample with the median incomes by household size for the U.S. population.

Table 1

U.S. Census Demographic Variables

Demographic Category	Percent of population
Race and Ethnicity	
White and Non-Hispanic	57.8
Black and Non-Hispanic	12.1
Asian and Non-Hispanic	5.9
Multiple Racial Categories and Non-Hispanic	0.76
White and Hispanic	18.7
Income	
Modal income range ($150,000 or more)	27.6
Education	
Modal education level (bachelor's degree)	21.8

Note. The proportion of adults reporting each demographic variable on the U.S. Census (race and ethnicity) and the American Community Survey (income and education).

Interim Discussion

This study aimed to amass a set of photographs for commonly investigated early object nouns. Using a survey hosted on Prolific, we collected a set of 1043 photographs across 40 different word categories. Interestingly, participants did not contribute photographs for all categories of words equally. The majority of caregivers provided images for many of the word categories, but there were several words (i.e., mouse, moon, and radio) that were provided by only a small number of caregivers. The relatively limited sample of images for these categories may reflect broader experiences with these word categories. Specifically, one possible explanation for differences in the number of images provided for particular categories is that infants may not regularly experience those, and thus, we have fewer referent examples for these object categories. Alternatively, these words may be typically experienced in copyrighted materials (e.g., books), and therefore, caregivers avoided providing these images per the survey instructions.

We collected images from a small sample of caregivers, but our participants were relatively diverse. Caregivers reported a range of incomes and education levels, and, in general, the socioeconomic statuses represented in the current study appeared to reflect the broader U.S. population. We focused on socioeconomic status as a particularly important demographic factor because it is a strong indicator of the resources available in a given household. By sampling families with a range of resources, we are more likely to collect images that reflect the kinds of objects (e.g., toys, furniture, foods) experienced by infants in the U.S. Therefore, the diversity in income and education levels included in our caregiver samples suggest that the referents included in the resultant image database will be representative of many U.S. households.

Study 1b

The images collected in Study 1a were provided by a relatively small, albeit socioeconomically representative, sample of families in the United States. Thus, the images may only reflect the referent experiences of a small set of infants. The goal of the database, however, is for the images to be used widely by developmental scientists with many different populations. Therefore, Study 1b assesses the extent to which the referents provided in Study 1a generalize to the experiences of other U.S. infants by collecting a set of norms from a new set of caregivers.

Images may vary on myriad dimensions. We identified three dimensions that we hypothesized to be relevant for developmental contexts and, therefore, would be useful to include in the database: name agreement, familiarity, and similarity (Brodeur et al., 2010, 2014;

Snodgrass & Vanderwart, 1980). Name agreement assesses the extent to which subjects agree on a label for an image. Experience quantifies how often participants report experiencing an image. Lastly, similarity measures the pairwise similarity between two images.

Name agreement (i.e., *H* value) quantifies the number of unique names provided and the number of participants that endorse each name. For example, if there are two names provided equally by participants for a given image, the *H* value would be 0.5. *H* values provide a measure of whether caregivers in general would use the same label when referring to a particular image. However, it does not indicate which label is the most common, only the extent to which participants agree on a given label for an image. This metric is likely to influence the strength of the association between a word and the image of interest. This is particularly relevant for developmental experiments such as looking-while-listening, as word recognition depends both on the visual (Arias-Trejo & Plunkett, 2010; Garrison et al., 2020; Pomper & Saffran, 2019) and auditory (Potter & Lew-Williams, 2023) stimuli selected. Thus, collecting name agreement as well as the modal name(s) will provide future database users with a measure to adjudicate between referents for a given label depending on their question of interest.

Experience (i.e., familiarity) may also influence word recognition in infants. Familiarity has been manipulated in several experimental paradigms with varying results (Casey et al., 2023; Garrison et al., 2020). For some nouns, infants' word recognition was not dependent on prior experience with the referent image. However, for other words, infants were unable to recognize researcher-selected images as instances of a word, despite the same words being reported as the very early learned. Thus, infants' experience with a referent is a critical component to assess in an image database.

Word recognition accuracy depends on the similarity between the target and distractor images. For example, both typically-developing infants and autistic toddlers are slower and less accurate at fixating a target image when the distractor image has overlapping perceptual features (Ellis Weismer et al., 2016). Similarly, Weaver and colleagues (2024) found a small boost in infants' word recognition when the target and distractor were perceptually dissimilar. Thus, a quantitative measure of similarity between stimuli would be extremely useful when developing experiments.

Similarity can be operationalized in several different ways. For example, two items may be rated as perceptually similar because they share color or shape features. Prior research has used adult ratings of similarity in which people select pairs of images that they think are most alike (Ellis Weismer et al., 2016). While conducting adult ratings of pairwise similarity may be useful in a small-scale experiment, it is much more resource-intensive to conduct pairwise comparisons for a large set of images. As an example, a set of 500 images would have a set of 124,750 unique comparisons to rate. Furthermore, it is unclear whether adults' ratings of similarity map onto the features of similarity to which infants attend. An alternative to defining similarity using adult ratings is to use distance metrics drawn from large vision models such as *DreamSim* (Fu et al., 2023). *DreamSim* generates predictions of perceptual distance between image pairs and has been shown to closely approximate human ratings of similarity. Rather than using pixel-by-pixel or patch-level similarity, this metric focuses on mid-level similarity, which better captures similarities in the semantic content of images. Weaver and colleagues (2024) used *DreamSim* as a similarity metric when investigating similarity between target and distractor images, demonstrating the feasibility of this similarity rating in infant research. By including pair-wise image similarity in an image database, users can select target and distractor images according to this metric, depending on the research question of interest.

The current study aimed to norm a large set of images collected from caregivers. We developed a survey to collect labels and experience ratings in a sample of caregivers of children under two years of age. We also generated pairwise similarity ratings. Using these norms, we identified referents for each category that are likely to be highly experienced by infants. We also examined whether norms differed by age or word category (as defined on the MacArthur-Bates Communicative Development Inventories).

Methods

Participants

Using Prolific, we recruited US caregivers of infants under 24 months of age (M=15.78) who reported that English was one of their primary languages. The final sample size included 358 caregivers who did not participate in Study 1a. Each participant rated a subset of 75 images (out of 964 total images), and we aimed to have 30 caregivers rate each image. Therefore, participants were collected iteratively until a total of 30 usable participants rated each image. An additional 161 caregivers participated but were excluded for having a child outside of the target age range (N=61) or failing to pass the attention checks (N=100).

The sample size was determined by considering prior database studies and the number of eligible participants available on Prolific that we could reasonably recruit. Prior studies norming photographs of objects for language experiments typically include between 30 and 152 participants (Adlington et al., 2009; Brodeur et al., 2014; Krautz & Keuleers, 2022). In these experiments, participants are tasked with rating the entire set of images. However, our participants only rated 75 images. Pilot testing revealed that participants could only feasibly name and rate 75 images within a 30-minute testing window. While the target sample of N=30 per image is on the lower end of sample sizes in prior literature, we were limited by our participants' completion rate. Thus, 30 participants rating each image was a compromise from the proposed 50 ratings per image.

Caregivers largely identified as White and not Hispanic (65.64%). 20.11% of the sample identified as Black and not Hispanic; 3.63% identified as Asian and not Hispanic; 3.07% identified as not Hispanic and identified with multiple racial categories. An additional 7.54% of the sample identified as Hispanic. The sample was also educated (74% had a bachelor's degree or higher) and upper-middle class. The most commonly reported income range was $100,000 to $149,000 (30.45% of the sample). Notably, the sample of caregivers in Study 1b was more affluent than the sample of caregivers included in Study 1a, but was more similar to the national averages.

Procedure

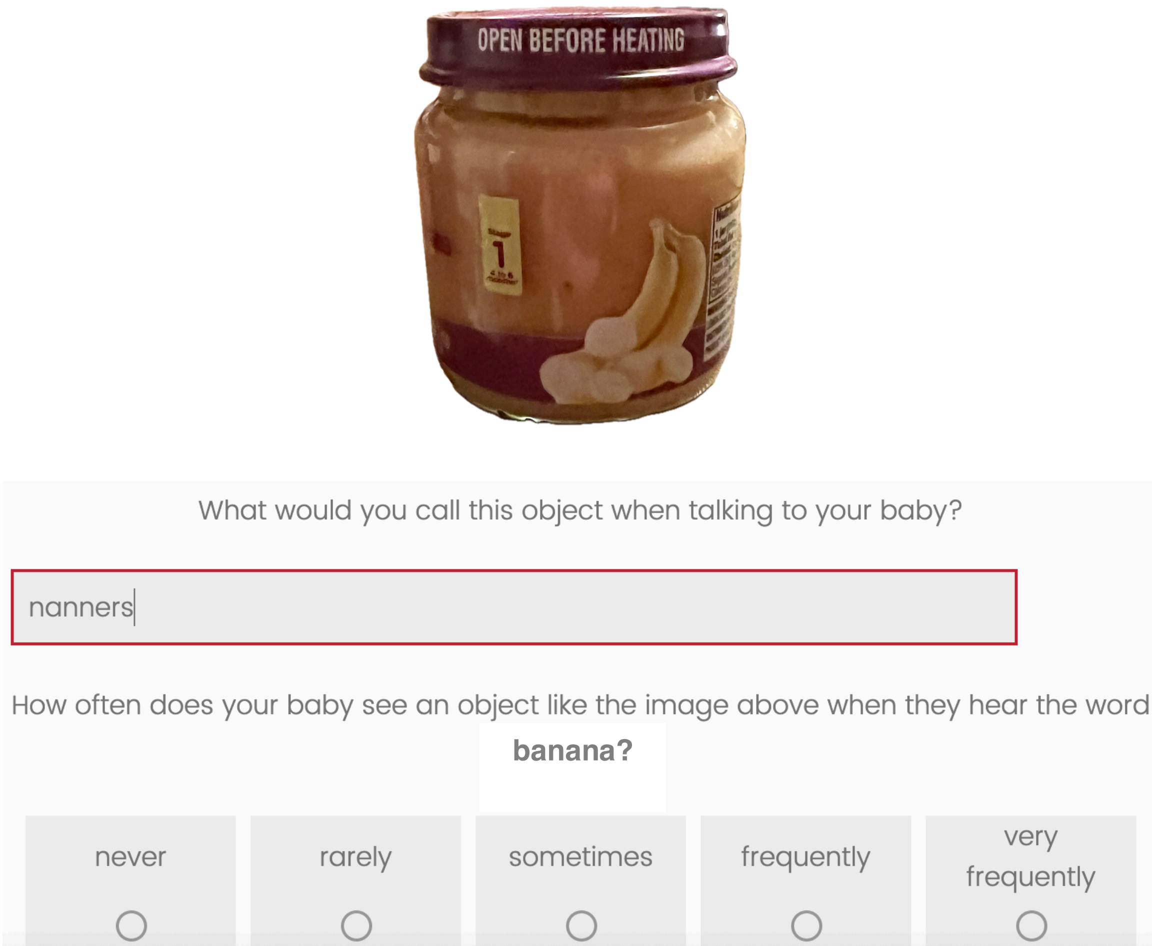
Prior to use in Study 1b, the raw images collected in Study 1a underwent several rounds of editing. First, any identifying information, including but not limited to faces, names, addresses, and license plates, was removed from the images. We followed HIPAA guidelines as well as guidelines outlined by researchers conducting home video and audio recordings (e.g., https://bergelsonlab.gitbook.io/blab/data-pipeline/personal-information-guidelines) to determine identifiable information. After deidentifying the photographs, the images were edited to isolate the target object and to improve lighting conditions using software including Apple Photos, Apple Preview, Adobe Photoshop (Adobe Inc., 2025), and GIMP (https://www.gimp.org/). The isolated versions of the images were saved as a separate copy from the original deidentified images so that the final database could include both versions of the images for future researchers who may be interested in studying objects in naturalistic contexts.

During the editing process, some images were deemed unsuitable for norming and removed from the database. Images that were blurry, too dark, or had an unidentifiable object (according to the author and a research assistant) were not normed. We reasoned that norming the stimuli was resource-intensive, and therefore, we focused on images that future researchers would likely use in their experiments. Ultimately, 964 images out of the total 1043 collected in Study 1a were retained for norming.

The norming survey took approximately 30 minutes and included a naming task and an experience rating task. Participants were presented with a random subset of 75 images (out of the 964 total images). For each image, they were asked to provide a noun that they would use when labeling the referent for their infant. They also rated how likely their infant is to see the referent image while hearing the target word, using a Likert scale from "never" to "very frequently". For instance, on a given trial, a participant might see an image of mashed bananas. They were first asked, "What would you call this object when talking to your baby?" After responding, they would see a second question that asked: "How often does your baby see an object like the image above when they hear the word "banana"*?*" The set of images was presented in a random order, but participants always named the image prior rating their child's experience with the image. See Figure 2 for an example trial.

Figure 2

Example trial for the norming survey



Note. A sample trial for the norming survey. Participants would be shown one image, and would provide the name for the object. Once they provide a name, they would be asked to report how often their child experiences the object when hearing the category label (i.e., noun reported on the MB-CDI).

Participants also completed demographic questions. They reported their race, ethnicity, annual household income, and education level. To ensure that participants had a child in the age range, they had to provide their child's age information on two separate occasions within the task. First, they reported their child's age in months at the beginning of the survey. After the 75-image rating task, they provided their child's date of birth. Any participants who reported an age for their child that did not match the date of birth they reported were excluded from analyses, as we were unable to verify their eligibility for the study.

In addition to the caregiver rating survey, we also collected a measure of similarity between each of the 964 referent images using *DreamSim* (Fu et al., 2023). Using *DreamSim*'s pairwise similarity function, we extracted perceptual similarity, which was defined as the cosine distance between their embeddings.

Measures

Name agreement. To assess name agreement, we calculated the proportion of participants who endorsed each name for a given image. These proportions were then used to compute the H value for each image, which is calculated using the formula

$$H = \sum_{i=1}^{k} p_i \, log(1/p_i),$$

where p_i is the proportion of participants who endorsed the ith name, and k is the total number of unique names. Therefore, name agreement captures the number of distinct names provided and the distribution of support for each name.

Experience. We defined two different measures of experience. For each image, we averaged participants' experience ratings (i.e., Likert scale from 1 to 5) to calculate the mean experience rating. We also calculated the modal frequency response for each image, which reflects the most commonly reported experience rating for a given image. We include both the mean and the modal experience value for this set of images because these two metrics may be capturing different aspects of experience. That is, for some images, there may be a wide variety of experiences endorsed by participants, which would result in a mean value somewhere in the middle. The modal value, however, may be relatively higher or lower depending on the number of participants reporting each experience level. Thus, we use both metrics to capture potential meaningful variability.

Similarity. Similarity was computed using *DreamSim*. The large vision model computes a vector representing each image. We then computed the cosine distance between each vector pair in the image set.

Results

The primary purpose of Study 1b was to develop a set of norms for the set of 964 images collected in Study 1a. Photographs of referents were provided for 40 different words; the number of photographs provided per word varied (Range = [10, 33]). Below, we reported descriptive statistics for the norms collected in Study 1b (i.e., name agreement, experience, and similarity) averaged across the entire sample of images and for individual word categories (Appendix B). In the final database, we will report the image, the image identifier (e.g., apple01.png), the number of participants who rated each image (*N*), the modal name, the *H* value, the mean experience, the standard deviation of experience, and the modal experience level. Additionally, a separate file will report pairwise distances for each image. A sample of the database can be seen in Appendix C.

Table 2

Summary Statistics for the Entire Image Set

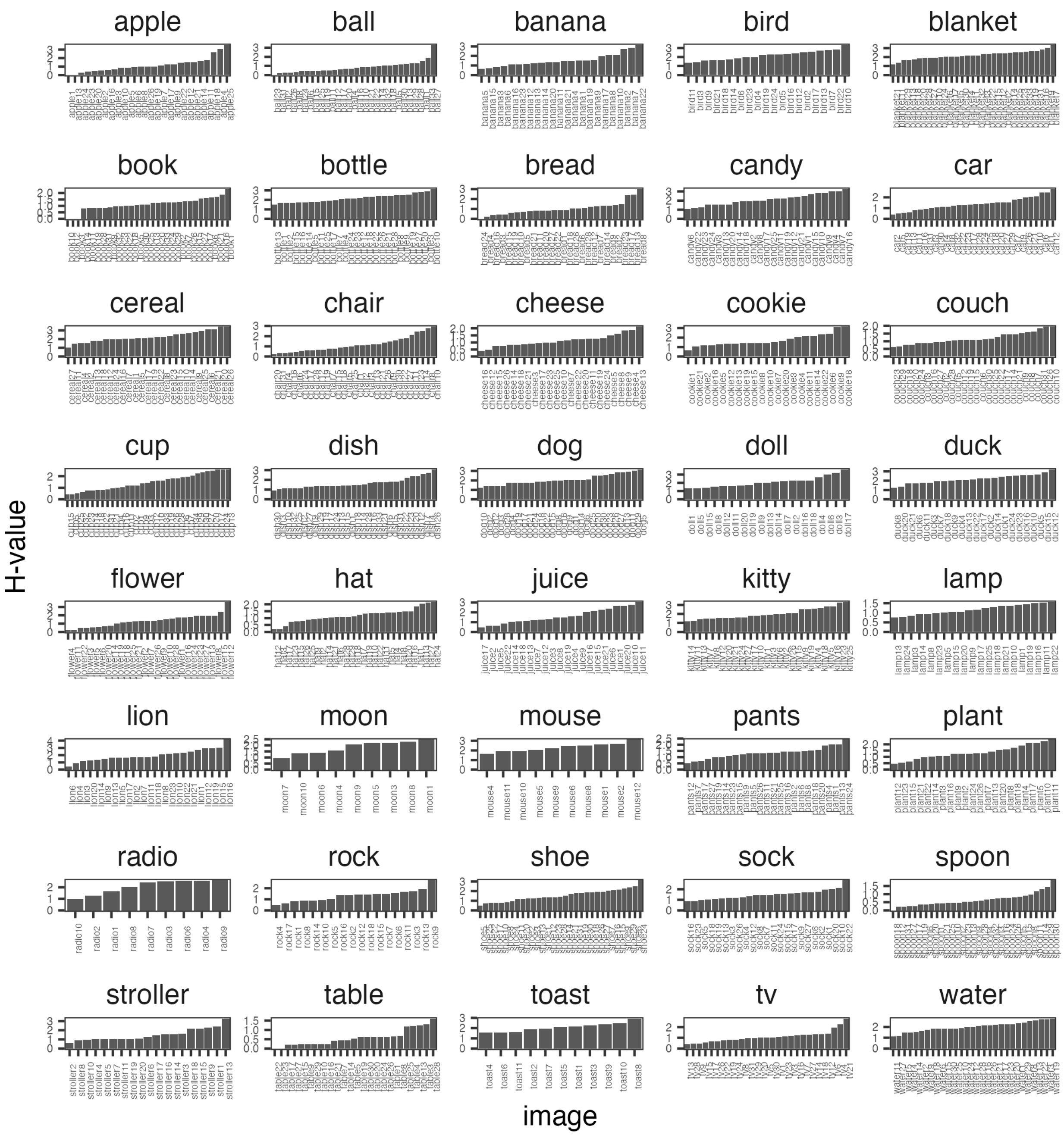
	H value	Experience Rating	Similarity
M	1.534	3.184	0.739
Mo		5	
SD	0.738	1.653	0.086
Range	0 – 4.131	1 – 5	0.096 – 0.988

Note. The *H* value, experience rating, and similarity ratings were averaged across all individual images. We report the mean (*M*), standard deviation (*SD*), and range of values. For experience rating, we also report the modal rating (Mo).

We first examined name agreement. On average, 6.58 names were provided for each image, and each name was endorsed by 15.19%. For the most common name provided (i.e., modal name), agreement reached 67.68%. The average *H* value (*H*=1.541; Table 2) was higher than in classic stimuli databases such as the Snodgrass and Vanderwart (1980) images (i.e., *H*=0.558), but is commensurate with databases using photographs (i.e., *H*=1.65; Brodeur et al., 2010, 2014). We also found high variability in name agreement across categories (see Appendix B for the average *H* value by word). As can be seen in Figure 3, the images in some categories (e.g., ball, spoon, table, and TV) have relatively low *H* values across the category, which indicates high name agreement for the images included in these categories. Other categories (e.g., blanket, cereal, and mouse) have high *H* values across the images, suggesting generally lower name agreement for these word categories.

Figure 3

H value for the image set



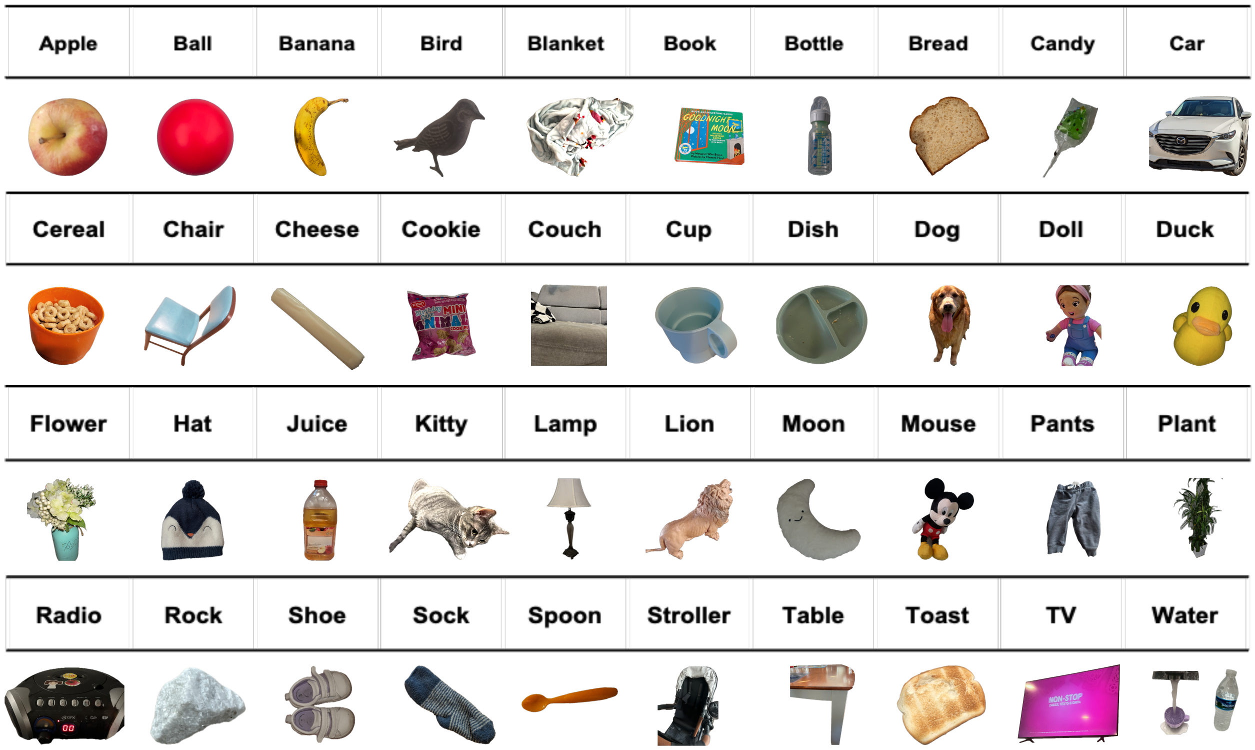
Note. *H* values for each image. Each panel groups the images by word category. The x-axis depicts each individual image for a word category. The y-axes have different values based on the range of *h* values for a given word (0 to 3).

Next, we investigated the experience ratings. On average, images were rated at a moderate level of experience (*M*=3.1; Table 2), which approximately corresponds to an experience rating of "sometimes" on the scale used in this study. The modal experience rating, however, was 5. This high number suggests that across the images, the majority of the sample reported that the images were very frequently experienced by their child. Notably, the modal experience rating varied from 1 to 5 across individual images, suggesting that while the majority of images were reported to be very highly experienced, other images were less representative of children's referent experiences for specific words. The modal ratings for specific images were endorsed by 41.01% participants, on average.

In addition to examining trends in experience with specific images, we investigated which images were reported as the most highly experienced for a given word. For each word, we selected the image that was rated as highly experienced by the largest proportion of participants. We selected the modal experience as the central tendency of interest for this measure because the mode reflects the image that was most commonly reported as very frequently experienced. In general, the images that were rated as very frequently experienced appear to be relatively stereotypical exemplars (Figure 4). However, the perspectives (i.e., viewpoints) of the photographs are much more varied than the kind of stimuli typically used in infant studies. For example, the image that was most frequently experienced for the word "apple" depicts a whole apple photographed from above (Figure 4) rather than the side view (e.g., Kartushina & Mayor, 2019).

Figure 4

Very Frequently Experienced Images



Note. This figure presents the images that were most commonly reported as very frequently experienced (i.e., mode) for a given word. Two modes existed for the word "water", and thus, both images are reported.

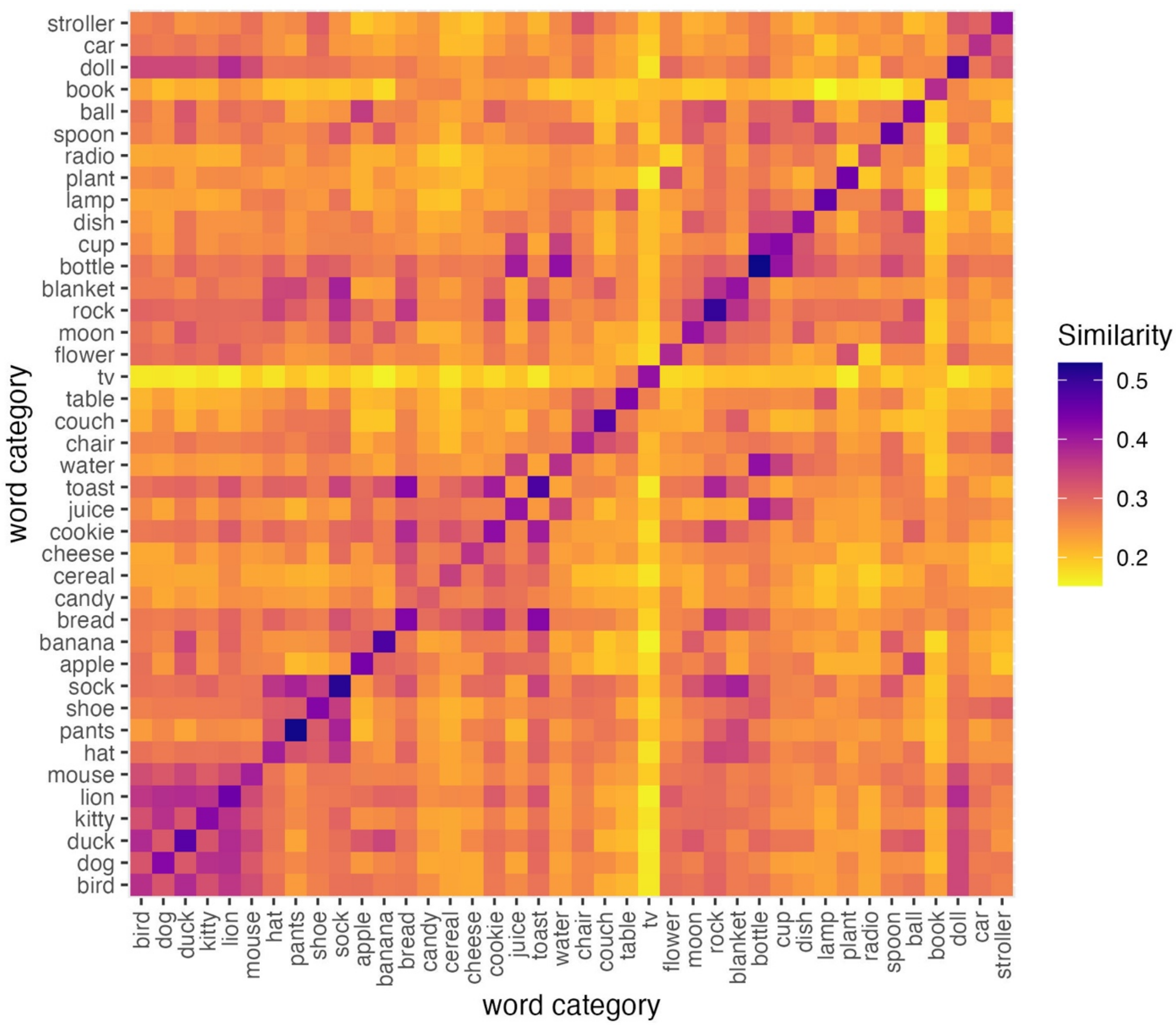
Name agreement scores and experience ratings varied immensely across the sample of images. The previous data, however, do not indicate whether these norms vary randomly or whether the two norms are related in meaningful ways. For example, there may be images that are reported as rarely experienced and have very little name agreement. Therefore, we investigated whether name agreement and experience were correlated. We fit a linear mixed-effects model regressing average experience per image on *H* value (cluster mean centered; Enders & Tofighi, 2007), including a by-item random slope and intercept for word. *H* value was a significant negative predictor of experience (b=-0.135; $F(1,33.555)$=32.017; p<0.05), suggesting that images with larger *H* values (i.e., less name agreement) were rated as less frequently experienced.

The final norm of interest was the pairwise similarity ratings computed using *DreamSim*. Across the entire sample, average similarity was high but varied across pairs (Table 2). To visualize overall patterns in similarity, we created a heat map of similarity scores (Figure 5). We averaged pairwise similarity scores at the word level. Specifically, we averaged similarity scores for comparisons between images within the same word and for comparisons between images for a given word and the images for each other word. For example, we calculated an average similarity score for all pairs of images for the word “apple”, and we calculated an average similarity score for pairs of images for the words “apple” and “banana”, “cheese”, etc. The heat map was also organized according to the categories included on the MB-CDI (see Appendix A for the full list) to investigate whether words from similar superordinate categories were rated as more similar to one another.

Visual inspection of Figure 5 suggests two general patterns in the similarity scores. First, words from the same MB-CDI category appear to be more similar to one another than to words from another MB-CDI category. For example, images for animal words are more similar to images for other animal words than to images for clothing words (Figure 5). Secondly, images from the same word category appear to be more similar on average than images from different word categories. That is, average similarity scores for comparisons along the diagonal of Figure 5 are much higher than similarity scores for comparisons off the diagonal.

Figure 5

Heat Map of Similarity



Note. This heat map depicts pairwise similarity ratings computed from *DreamSim*. Each cell includes the similarity ratings averaged across all images for each pairwise comparison. Darker (purple) colors indicate pairs of words that have a higher average similarity, and lighter (yellow) colors indicate lower average similarity. Axes were ordered according to the MB-CDI categories.

We further investigated the patterns of similarity for within and between word categories using a modeling approach. For this analysis, we calculated two similarity scores for each image. One score reflected the average similarity for comparisons for a given image and all other images provided for the same word (e.g., within word category). The second score reflected the average similarity for comparisons for a given image to all images from other word categories (i.e., between word categories). We then fit a linear mixed-effects model in which we predicted similarity score from comparison type (coded -0.5 for between word categories and 0.5 for within word category), including a by-item random intercept. The effect of comparison type was significant (b=0.169, $F(1,1004.4)$=4965.1, p<0.05) such that pairwise similarity was higher for images within the same word category than for images between different word categories. These results confirm the pattern of similarity observed in Figure 5.

Interim Discussion

In Study 1b, we edited and normed a set of 964 images for a referent database. Caregivers provided the name that they would use when labeling the image for their child and rated the frequency with which their child experienced each image. We found that name agreement varied widely across the sample of images, as did experience with each item. This variability, however, was not random. There was a negative relation between name agreement and experience, suggesting that images with lower overall name agreement were rated as less frequently experienced.

Although we did not predict relations between the two norms, the pattern of the current results is commensurate with prior work on image databases. Several databases have reported negative, moderate correlations between name agreement and familiarity (Adlington et al., 2009; Berman et al., 1989; Brodeur et al., 2014). The negative relation in the current study may be well explained by the fact that caregivers had limited prior experience labeling objects that were not commonly encountered by their child. Thus, they may have reported more varied names across participants. For example, some caregivers may have used more adult-appropriate labels, while other caregivers may have invented a child-appropriate version of the typical name for the referent. In general, the negative relation indicates a lack of name agreement amongst referents that are rarely encountered, possibly indicating that those kinds of referents are not well-suited for use in experimental settings.

In addition to the caregiver-reported norms, we computed pairwise similarities for the entire image set using *DreamSim* (Fu et al., 2023). We found that images that belonged to the same word category were rated as more similar than pairs of images that belonged to different word categories. This result reveals that there are features shared amongst referents for a given word that are distinctive from other word categories. While the existence of shared features within a category have been well established in prior research on categories (e.g., exemplar theory and prototype theories all include some level of featural similarity; Mervis & Rosch, 1981; Rosch et al., 1976; Rosch & Mervis, 1975), this study provides preliminary evidence for similarities within real-world object photographs, many of which varied highly across referents. For example, photographs of apple slices, apple puree in a baby food jar, whole red apples, and applesauce pouches were all provided as referents for “ apple”. Yet, these referents were still rated as more similar to one another than to referents from other words. This suggests that there are sufficient perceptual similarities between these diverse referents for the large vision model to be able to extract meaningful within and between category relations.

The results of the similarity data also provide evidence that large vision models may be particularly useful in developing stimulus norms. Only one prior study, to our knowledge, has leveraged *DreamSim* to examine stimulus effects in infancy (Weaver et al., 2024). We provide additional evidence that similarity ratings drawn from this model may be a quick and easy approximation of human expectations. For example, we would expect human adults to also rate visually dissimilar referents of apples as more similar to one another than to referents of dogs because all apple referents are essentially just transformations of the same object. While this knowledge may be largely driven by adult knowledge that the same food object can be transformed into many different forms, it must also be somewhat related to shared perceptual features across the forms. *DreamSim* also seems to capture this relevant object information, as the differences in similarity for within-category comparisons and between-category comparisons cannot simply be explained by color or shape features alone. Future work could further validate this tool by comparing human ratings to the similarity metrics drawn from *DreamSim*.

One often overlooked aspect of prior databases is the demographics of the norming sample, which makes assessing the external validity of the databases a challenge. Therefore, we note that the present sample of U.S.-based caregivers who are educated and wealthy is a limitation of our study. In particular, the referent experiences reported in our norms likely will not reflect infants outside of the United States. Despite our efforts to recruit a relatively representative sample of participants, the resultant norms may also be less characteristic of the referent experiences for infants from lower socioeconomic backgrounds. One important future direction is to collect norms from caregivers from lower socioeconomic statuses.

General Discussion

This study collected and normed over 950 images that reflect infants' real-world experiences with a variety of referents for a set of frequently-tested nouns. While many photorealistic databases are widely available (Brodeur et al., 2010, 2014; Hebart et al., 2019; Krautz & Keuleers, 2022), this is the first planned database of real-world objects curated specifically for infant populations. In particular, we used a novel crowdsourcing approach to amass images that infants experience, directly from caregivers. This image collection strategy resulted in a highly variable set of referents for 40 different early learned words. Critically, by collecting images from caregivers, we avoided imposing our prior expectations about the kinds of referents infants experience on the resultant database. Thus, all the images included in the final data set are those that (at least some) caregivers endorsed as being relevant for infants' early experiences with words.

We also developed a set of norms for the database, which will be essential for future research using the database. As an initial set of norms, we included name agreement, experience, and similarity as each of these factors has been shown to influence language processing in prior work (Arias-Trejo & Plunkett, 2010; Garrison et al., 2020; Moore & Bergelson, 2024; Potter & Lew-Williams, 2023). The norms varied across the sample of images, but were, on average, similar to other adult image databases. These results were quite surprising as most prior databases typically norm high-quality, professional photographs. Yet, our images were photographed by caregivers in their homes, most likely using cell phone cameras. Despite our unconventional image collection approach, we were able to develop a high-quality database with valid norms.

The norms are an essential aspect of a database as they allow researchers to control for confounding variables. For instance, Casey and colleagues (2023) were unable to select referents that represent infants' experiences with some early-learned words. The current database could ameliorate this challenge, at least for the set of words included. That is, researchers can easily select stimuli that are reported as highly experienced (using either the mean or the mode values provided). Similarly, prior studies demonstrate strong evidence that word recognition is slower when the target and distractor images are similar (Arias-Trejo & Plunkett, 2010; Bergelson & Aslin, 2017; Ellis Weismer et al., 2016). This factor can be controlled for in future studies by including images that have low pairwise similarity. Moreover, the norms could be used to increase or decrease the difficulty of a given experimental paradigm very easily. For example, a researcher could initially select pairs of stimuli that are rated as highly dissimilar to use for piloting a task. If many children are at ceiling or there is not sufficient variability, the task could be made more challenging by selecting different referents that have higher pairwise similarity ratings or that are more infrequently experienced. Taken together, these examples highlight the utility of these norms for designing and refining experimental tasks.

The database also provides new opportunities to experimentally manipulate variables related to the norms. For example, future studies could examine the extent to which name agreement impedes language processing, which would build on the literature demonstrating that specific labels influence word comprehension (Potter & Lew-Williams, 2023; Swingley & Aslin, 2000). An experiment could include referents that vary in name agreement or include various modal names for a specific word category. The similarity norms could also be leveraged to examine word recognition. For instance, a future study could investigate typicality effects in categories of words that have low pairwise similarity. That is, typicality effects may be stronger for words that have relatively low perceptual similarity, as it may make identifying distinctive shared features more challenging. These questions, as well as many more, can be easily investigated using our new database norms.

More broadly, this experiment represents the first large-scale investigation of 0- to 24-month-olds' referent experiences with the nouns on the MB-CDI. Understanding the range and kind of referents experienced by infants allows future studies to more accurately probe early word knowledge. For example, we can identify categories of words that have relatively more variable referents using this new database, which may be particularly important for examining individual differences in early word comprehension. Specifically, we can interrogate word categories that were reported to have a wide range of experiences across various referents or word categories that had relatively lower pairwise similarity ratings between referents. While both kinds of referent variability may have important implications for infants' word knowledge, they have been relatively understudied, potentially due to the lack of readily available information about the kinds of referents experienced. The present database, however, is an initial step to begin addressing these gaps.

Limitations

Although we specifically selected norms based on prior language research, the current database is limited in the number and kinds of norms included. In particular, we did not include a measure of visual complexity, which is commonly normed in other databases (Adlington et al., 2009; Berman et al., 1989; Brodeur et al., 2010, 2014; Hebart et al., 2019; Snodgrass & Vanderwart, 1980). Visual complexity is typically defined as a subjective rating of how much detail or intricate lines are included in a given image. While visual complexity may influence language processing in infants, there is no evidence that visual complexity, as defined by an adult caregiver, would reflect infants' perception of visual complexity. Future studies could measure visual complexity with infant populations, perhaps using a preference paradigm (e.g., Karmel, 1969).

Another important limitation of the current database is the relatively small, U.S.-based sample used for image collection and norming. While our sample in Study 1a was diverse with respect to U.S. socioeconomic statuses, it is very possible that the range of referents included in the current database is not representative of the broader population's experiences. The referent experiences in the current studies are also unlikely to generalize to other cultures and/or countries outside of the United States. While the experience norms provide evidence that some of the referents were very frequently experienced by a separate set of infants in the U.S., they do not directly indicate that a given referent is the *most* commonly experienced referent for a given participant. That is, a caregiver may have reported multiple referents as very frequently experienced, but we did not measure which (if any) referent was most typical of their child's experiences with a given word. We could examine additional norms using the current database with broader caregiver samples in future studies to better understand infants' referent experiences.

Finally, the particular participant pool is important to consider for future work. Prolific allowed us to recruit participants who were more diverse than typically included in studies of US infants (Nielsen et al., 2017; Roberts et al., 2020), though the norming sample was less socioeconomically diverse. However, because we recruited participants to take part in an unmoderated survey, we were unable to verify that these participants actually had a child. We included a validity check, which did result in a high proportion of unusable participants, but some of the participants included in the database may not have been caregivers. Thus, it would be important to validate these norms in person with participants with whom we can confirm the age (and existence) of their child. While an in-person study would undoubtedly be less representative of the broader population, it would provide additional confirmation that the online results were valid.

Conclusion

Ultimately, this database will make a significant contribution to the early language community. The images can be used to generate new questions about the nature of word-referent associations across different word categories. Moreover, this database will improve experimental design by providing a set of normative images, potentially reducing researchers' degrees of freedom in the kinds of stimuli used to examine familiar word knowledge. Due to the database's widespread application, this database has the potential to be an extremely valuable methodological tool for developmental scientists.

Study 2

Mealtimes are rife with opportunities to learn about the meaning of words. Caregivers and their infants engage in rich social interactions, including touching, looking at, and labeling foods (Custode & Tamis-LeMonda, 2020), all of which are embedded in a specific physical location in the home (i.e., the kitchen). Prior research has identified both social interaction and unique physical contexts as strong predictors of early word learning and vocabulary growth (Breitfeld & Saffran, 2024; Roy et al., 2015; Suarez-Rivera et al., 2022; Tamis-LeMonda et al., 2014; Tamis-Lemonda et al., 2018). Recent evidence also suggests that the visual and labeling statistics present during mealtimes may facilitate word learning (Clerkin et al., 2017; Clerkin & Smith, 2022; Slone et al., 2023). There are many objects frequently present at mealtimes, potentially helping to reduce referential ambiguity (Clerkin et al., 2017). While caregivers name very few objects during mealtimes, they often label objects that are present in the infants' field of view (Clerkin & Smith, 2022). Furthermore, mealtimes occur frequently throughout the day, offering multiple opportunities for infants to engage in precisely the kinds of activities that foster word learning (Clerkin et al., 2017; Clerkin & Smith, 2022; Custode & Tamis-LeMonda, 2020; Tamis-Lemonda et al., 2018). Therefore, it is no surprise that food words are among the earliest learned words in development (Bergelson & Swingley, 2012; Roy et al., 2015).

Naturalistic recordings of infants' visual experiences have uncovered two interesting findings about early-learned words. First, infants likely build a rich representation of an object referent before learning its label (Clerkin et al., 2017; Clerkin & Smith, 2022). Secondly, infants may have idiosyncratic experiences with early words, but researchers do not typically test these variable referents in experiments that measure early word comprehension. Indeed, Clerkin and Smith (2022) found that the word “banana” almost always co-occurred with mashed or cut-up bananas, but not with whole bananas during 9-month-olds’ meals. Despite this observation, most LWL studies testing “banana” use canonical images of whole bananas in their peels. Casey and colleagues (2023) similarly found that abstract words (e.g., uh-oh) were said in the presence of several common referents (e.g., a broken crayon for “uh-oh”) that rarely matched the experimenter-selected referents for an LWL task (e.g., a cup falling for “uh-oh”). It is, therefore, possible that 10- to 16-month-olds failed to recognize these early learned words in a looking-while-listening paradigm because of the mismatch between infants’ lived word-referent experience and the word-referent pairs tested in the task. These findings suggest that infants’ early experience with referents lays the foundation for understanding the meaning of words.

Prior research examining object experience in word learning has focused on understanding the underlying visual statistics as a mechanism for solving referential ambiguity; however, infants are also learning about the specific referents that co-occur with the labels they hear. For example, if infants frequently experience cut-up pieces of banana, then they may be developing a rich representation of this form of banana as the likely referent for the word “banana.” Adults, however, may typically experience whole (often peeled) bananas and, as a result, consider the word “banana” as referring to a whole banana. This example suggests that referent experiences may not be consistent across individual infants or developmental time, resulting in qualitatively different word representations. Yet, there is limited prior research that examines how prior experience with specific referents influences infants’ word representations.

Existing work has focused on familiarity or typicality to investigate whether word comprehension is facilitated by certain kinds of referents. For example, one study interrogated 12- to 18-month-olds' word comprehension in the presence of highly familiar referents (i.e., an image of a Croc shoe they own at home) and in the presence of prototypical unfamiliar referents (i.e., another infant's shoe; Garrison et al., 2020). Although there was a slight, non-significant boost in word recognition for familiar referents for 12-month-olds, older infants' word comprehension was equally accurate across both familiar and unfamiliar referents, suggesting that prior familiarity with an object has only a small effect on word meanings. Studies examining referent typicality similarly find very small effects of referent experience (Weaver et al., 2024). Indeed, 14-to-18-month-olds robustly recognized atypical (e.g., kookaburra) and typical category members (e.g., robin) regardless of whether these referents were representative of infants' own category experience (Weaver et al., 2024). These studies suggest that early word representations may not vary as a function of prior referent experiences as indexed by familiarity or typicality; however, referent experiences may vary more substantially for some kinds of words than others.

Prior studies may have been limited in their ability to detect the relationship between experience and word representations due to the focus on shape-based words. For example, the referents for categories such as shoes and dogs have characteristic features that distinguish them from perceptually adjacent categories (socks and cats, for example). Food words may provide a stronger test of experience-based individual differences in word meanings than shape-based categories because infants encounter foods in diverse forms—whole, mashed, or cut—offering variable input (Clerkin & Smith, 2022; Perry et al., 2014). Whole bananas, cut-up bananas, and mashed bananas do not share the same shape, but are rather transformations of the same fruit, potentially leading to more idiosyncratic meanings for the word "banana".

Food words are also of interest because infants' experiences with foods diverge greatly both across age and across families. Caregivers may first introduce solid food in pureed forms because infants without teeth can easily swallow this texture. After infants' first birthdays, caregivers often let their infant eat cut-up pieces of food because they typically have teeth to adequately chew food and the manual dexterity to feed themselves. However, there is also variability between families in how solids are introduced, including baby-led weaning and spoon-feeding. Caregivers who use baby-led weaning serve small pieces of soft food that infants can easily pick up and chew, skipping pureed foods entirely. Caregivers who use spoon-feeding, however, use a spoon to feed their infant pureed or mashed foods until the infant develops some teeth and a more accurate pincer grasp (i.e., 9- to 12-months). This natural variability in the kinds of foods that infants may experience suggests that prior referent experience may exert a larger influence on infants' understanding of food words relative to other categories of words.

As such, the present study investigates whether and how infants' food word representations are shaped by experiences with different food referents. We employed two measures of experience to better capture individual differences and age-related differences in referent experiences. Infants' individual referent experiences were measured using caregiver report, while group-level differences were tested through the inclusion of two discontinuous age ranges. We predicted that younger infants would have qualitatively different referent experiences than older infants. Finally, we measured infants' word representations with a looking-while-listening (LWL) paradigm (Fernald et al., 2008) using two different kinds of referents. For each food word tested, we selected a form of the food that may be more common for younger infants (e.g., pureed or mashed) and one that may be more common for older infants (e.g., cut-up pieces). We hypothesized that infants would be more accurate in recognizing words in the presence of referents that reflect their lived experiences. That is, infants should be more accurate in recognizing referents reported to be more frequently experienced than referents reported to be less frequently experienced. Similarly, we expected group-level differences in recognition for the two kinds of referents, such that younger infants would be more accurate for younger-appropriate referents and older infants would be more accurate at older-appropriate referents.

Methods

Participants

Seventy-eight infants (N=45 females) and their primary caregivers participated in person at the Waisman Center. Two age ranges of infants were recruited, a younger range from 9- to 11-months (N=44) and an older range from 15- to 17-months (N=34). Infants were full term (born no more than 28 days early) and typically developing (per parental report) with normal/corrected vision and hearing. The participants were primarily white (N=67), upper-middle to upper class (N=51 reported an annual income greater than $100,000), and highly educated (N=66 reported holding a bachelor's degree or higher). Infants varied in their exposure to English (M=94%; Range: [10%, 100%]), but, importantly, their caregivers reported using the English labels most often when speaking to their child about the object. An additional 25 infants participated but were excluded for fussiness (N=10) or failing to contribute usable looking data for at least two trials for each referent (excluding the canonical forms; N=15).

The target sample size was determined using an *a prior* power analysis. We predicted that the effect of experience on word meanings would be small (i.e., $d = 0.28$), reflecting similar effects found in the literature (Garrison et al., 2020). A power analysis conducted using G*Power v.3.1 indicated that a sample size of $N = 82$ was required to achieve 80% power to detect an effect as small as those reported in the literature for the experience effect of interest.

We selected the current age ranges based on two primary considerations. First, we wanted to increase the likelihood that our sample would have variable experience with the target referents. We expected the younger group of infants would have more experience with food in very soft forms (e.g., applesauce) due to the recent introduction of solid foods at 6 months, while the older infants would have more experience with larger chunks of food (e.g., apple slices) because they typically have more teeth, potentially reducing choking risks. By including two discontinuous age groups, we can examine group-level differences in referent experience to understand whether age-related changes in the forms of food served to children influence the representations of food words. Secondly, we wanted to strike a balance between above-chance word recognition and evidence of individual differences in word representations. Recent evidence suggests that recognition of food words in some populations of 6- to 9-month-olds depends on the specific stimuli tested (Kartushina & Mayor, 2019). Even in studies demonstrating significantly above-chance word recognition in 6- to 9-month-olds, the effects are quite small (Bergelson & Swingley, 2012). However, infants older than 12 months have demonstrated robust word recognition in the presence of many referents regardless of their prior experience with the stimuli (Garrison et al., 2020; Weaver et al., 2024). Therefore, we included both a younger age range and an older age range to try to ameliorate these competing considerations.

Stimuli

Object words. Three foods were selected as the target words (i.e., *apple*, *banana*, and *cheese*). We only included three words to ensure that we could test each referent multiple times to assess individual differences at the word level and to ensure that we could maximally pair the images to reduce saliency effects. These specific words were selected because they are highly frequent in early caregiver-infant interactions (MacWhinney, 2000) and have a relatively early age of acquisition according to WordBank norms (see Table 3; Frank et al., 2017). Furthermore, prior research with 9-month-olds suggests that younger infants recognize the words “apple” and “banana”. Perhaps most important, however, is that these foods all have referents that can appear in many different forms. For example, apples can be served to infants as a pureed sauce or cut-up slices.

Table 3

Word Properties

Word	WordBank AOA (Comprehension, in months)	CHILDES (word frequency, ppm)
apple	15	1938.559
banana	12	2117.214
cheese	14	2328.697

Note: The AOAs reported are for American English. AOAs are defined as the point at which 50% of infants are reported to comprehend a given word, according to a bank of MB-CDI data. The prediction is based on fitting a logistic regression curve to each item’s data in WordBank using the fit_aoa() function from the wordbankr package. The CHILDES word frequency was calculated using recommendations from Sanchez et al. (2019) using the childesr package. The value is reported in parts per million (ppm) based on the number of word occurrences heard as input to 9- to 11-month-olds in the American English corpus of CHILDES (Braginsky et al., 2019)

Images. Two kinds of referents were selected for each food noun: a younger-appropriate image and an older-appropriate image. We also included one canonical form of each item, which was not central to the main hypotheses but allowed for exploratory examinations of typically-used referents (Figure 6). The images were generated using Gemini AI (Google, 2024) and created to conform with the age-related recommendations from the Centers for Disease Control and Prevention, the American Academy of Pediatrics, and parenting blogs aimed at early food introductions (e.g., https://solidstarts.com/foods/ and https://kidseatincolor.com/).

Auditory stimuli. A female native speaker of the local dialect recorded auditory stimuli using infant-directed speech. Each sentence included a directive carrier phrase followed by a target noun. Four different sentences that are commonly used with infants across both age ranges were included (Bergelson & Swingley, 2012; Campbell & Hall, 2022; Garrison et al., 2020; Weaver et al., 2024): “Look at the [target label]!”, “Can you find the [target label]?”, “Do you see the [target label]?”, and “Where’s the [target label]?”. The sentences were edited to be the same length and amplitude (65dB) with the onset of the target word occurring after 2,903 milliseconds.

Procedure

Infants completed an LWL task that tested their comprehension of the three target nouns. They were seated on their caregiver’s lap in a soundproof booth equipped with a 55-inch LCD television and video camera. A Tobii Spark eye tracker recorded eye gaze at a sampling rate of 60 Hz. Each session was also video recorded for manual eye-gaze coding if the eye tracker failed to maintain an adequate track. Before the LWL task, participants completed a five-point calibration procedure.

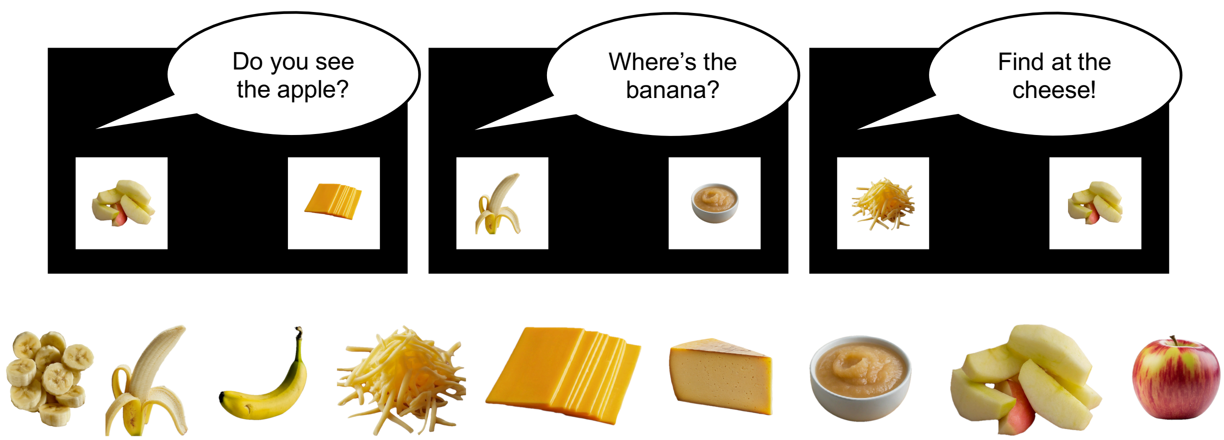
Following calibration, infants were presented with 32 test trials. The first two trials did not test the target words but, instead, served as a warm-up to familiarize infants with the task. The subsequent 30 trials tested infants’ comprehension of the target words. Each word was tested 10 times: four times in the presence of the younger-appropriate referent, four times in the presence of the older-appropriate referent, and twice in the presence of the canonical form

(Figure 6). On each trial, infants saw two images on white backgrounds on either side of the screen and heard a sentence labeling one of the images. The trials were blocked such that infants were first tested on all of the younger and older referents, which were paired exhaustively (e.g., applesauce was presented as a target with sliced bananas, a whole banana, shredded cheese, and cheese slices) to minimize effects of saliency (Pomper & Saffran, 2019) and mutual exclusivity (Markman, 1990; Markman & Wachtel, 1988). The canonical referents were all tested in the final block of the experiment, and only co-occurred with other canonical forms (e.g., a whole apple was presented with a banana in its peel and a wedge of cheese).

Trial order was pseudorandomized within block to ensure that the same target word was not tested on more than two consecutive trials and that the same target referent did not appear sequentially. Target images appeared equally often on the left and right sides of the screen, and the target did not appear on the same side for more than two consecutive trials.

Figure 6

Stimuli



Note. One block of experimental trials and the stimuli for Study 2. The top panel demonstrates test trials for the older and younger referents. Infants saw referent images from two different word categories on a screen and heard a phrase including a noun. The bottom panel illustrates the nine referents for the experiment. For each word, one referent appropriate for younger infants, one appropriate for older infants, and one canonical referent were included.

In addition to the experimental task, we assessed infants' experience with the forms of each food. Caregivers saw an image of one of the referents included in the LWL experiment and were asked to provide the top three words that they would use to label the image when talking to their child. Any child whose caregiver did not include an English label for the target images was excluded from analyses (N=2). Caregivers also reported how often their child experienced each form of food on a scale from never (1) to very frequently (5) in the last week, the last month, and across their child's lifetime. For instance, a caregiver would see a picture of sliced bananas and be asked, "Think about the foods your baby saw in the last week. How often did your baby see a food that looked like this?" Caregivers also completed the MB-CDI Words and Gestures Short Form to determine whether infants are reported to comprehend the target words as well as other words. In addition to the primary measures of experience, we queried caregivers about their infant's eating habits, method for solid food introduction, and demographic information such as their education level and household income.

Measures

Gaze Behavior. Gaze behavior was captured by an automatic eye tracker for 54 participants, and manually coded in 33-millisecond (ms) bins using Peyecoder (Olson et al., 2020) for the remaining 24 participants due to poor automatic tracking. Seventeen percent of the hand-coded data was independently recoded by a second coder. Coders agreed on the gaze location for 95.06% of frames and agreed on 93.66% gaze shifts within one frame. Gaze behavior from the automatic eye tracker was down sampled so that fixations occurred in 33-ms bins to be comparable to the hand coded data.

Gaze behavior was used to assess infants' word recognition accuracy. For each trial, we calculated the proportion of fixations to the target image out of the total fixations to either the target or distractor image during a critical window (300 ms to 2800 ms) after the onset of the noun (i.e., word recognition accuracy). The length of time for critical windows of analysis varies in the literature with shorter windows from 300 ms to 1800 ms or 36 7ms to 2000 ms following target word onset for older infants (Garrison et al., 2020; Marchman & Fernald, 2008b; Swingley et al., 1999; Swingley & Aslin, 2000) and longer windows extending to 3500 ms post target word onset for younger infants (i.e., 6- to 9-months; (Bergelson & Swingley, 2012; Kartushina & Mayor, 2019). Given the inclusion of both age ranges, we selected a slightly longer critical window to capture reliable looks to the target image across both age groups, as done in a prior study with a large age range (Weaver et al., 2024).

The final data set included 1893 trials from 78 infants. An additional 504 trials were excluded from the analyses because the trials did not have enough usable data during the critical window (i.e., fixations to either image for at least 50% of the 33-ms frames). These excluded trials typically reflect instances of inattentiveness (i.e., off-screen looking) during the majority of the trial.

Experience. Referent experience was defined in two ways in this study. First, we categorized the referents into two item types. Referents that were more likely to be experienced by younger infants (i.e., applesauce, banana slices, and shredded cheese) were assigned to the younger image category, and the remaining images (i.e., apple slices, peeled banana, and cheese slices) were assigned to the older image category. The canonical referents were excluded from the present analyses as they were not central to the research questions. Future exploratory analyses will examine word recognition in the presence of the canonical referents.

The second measure of experience was caregivers' reports of their child's exposure to the referents in the past week. We use recent experience as our primary measure, as adult work points to recent experience as strongly influencing word meanings (Rodd et al., 2016).

The methods, exclusionary criteria, and analytic plan were all preregistered on the Open Science Framework (https://osf.io/tahpk).

Results

Modeling Approach

The data processing and analyses were conducted in R (Version 4.5.0; R Core Team, 2024). Linear mixed-effects models (LMEM) were fit using the lme4 package (Version 1.1.37; Bates et al., 2015), and *p*-values were computed using Anova() from the car package (Version 3.1.3). This function uses Kenward-Roger approximation to estimate degrees of freedom for *F*-tests. Models were fit with the maximal random effects structures (Brauer & Curtin, 2018), and were iteratively pruned until convergence was reached following recommendations from Brauer and Curtin (2018) and Muradoglu and colleagues (2023). Specifically, we began by removing covariances between random effects, followed by removing random slopes for covariates and/or random intercepts. If all other methods fail to reach convergence, we retained random intercepts and removed random slopes for effects of interest as a final potential solution.

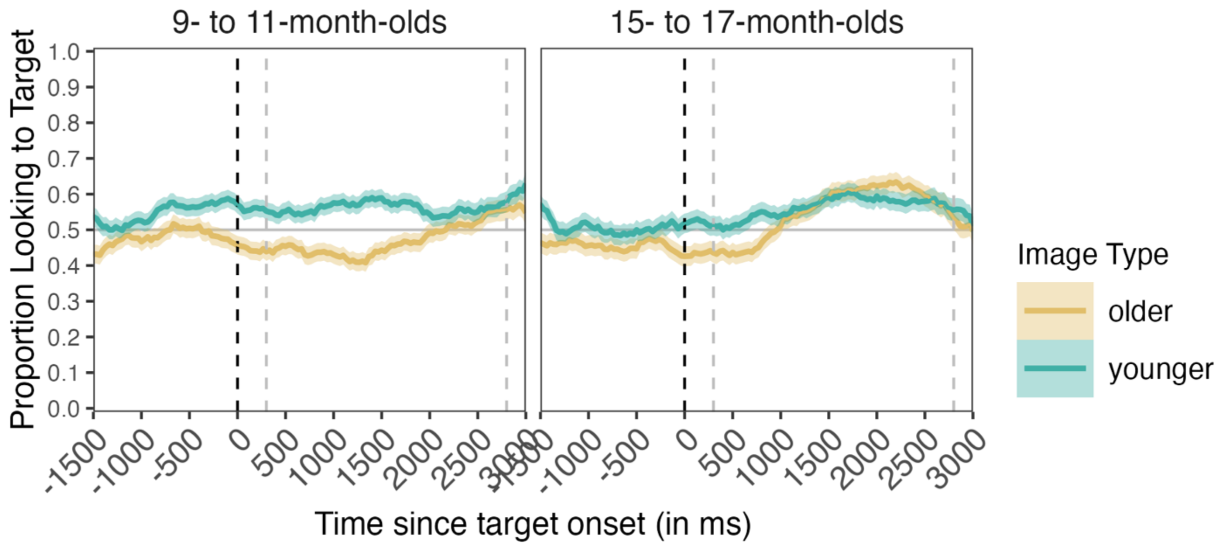
Analyses

Our main question of interest is whether experience with referents shapes word meanings, as indexed by word recognition in the presence of those referents. The first set of analyses operationalized experience using the two different image types (e.g., older vs. younger). We hypothesized that if word recognition is influenced by experience with referents, then word comprehension should be superior for age-appropriate referents. Following our preregistered analysis, we fit an LMEM regressing accuracy on image type (coded -0.5 for the younger image category and 0.5 for the older image category), age group (coded -0.5 for younger infants and 0.5 for older infants), and their interaction, including a by-subject random intercept. On average, infants recognized words above chance ($b = 0.033$, $F(1,73.62)=16.577$, $p=0.0001$). The effect of image type ($b = -0.053$, $F(1,1515.01)=10.715$, $p=0.001$), age ($b = 0.046$, $F(1,73.62)=8.065$, $p=0.006$), and their interaction ($b = 0.086$, $F(1, 1515.01)=7.017$, $p=0.008$) were all significant predictors of infants' word recognition accuracy.

Visual inspection of Figure 7 suggested that there were baseline preferences for the target images, especially for the younger infants. Thus, we wanted to assess the robustness of the effect of image type by fitting an analogous model using an alternative outcome measure for accuracy that corrects for baseline preferences. We calculated this alternative measure of accuracy by subtracting the proportion of fixations to the target image during a baseline window (-1667 to 0 milliseconds) before the onset of the target words from the proportion of target fixations during the critical window. This method has been used in prior studies to correct for infants' preferences for certain stimuli (Bergelson & Swingley, 2012; Weaver et al., 2024). After controlling for baseline preferences, the effect of age was the only significant predictor of accuracy ($b = 0.081$, $F(1,1388.13)=17.544$, $p<0.05$), with older infants outperforming younger infants. Therefore, the effect of image type and the interaction effect between image type and age were not robust to different measures of accuracy, suggesting that baseline preferences for the images from the younger category may have been driving the effect of image type in the preregistered analysis.

Figure 7

Word Recognition by Image Type and Age



Note. Time course of gaze behavior for younger infants (9- to 11-month-olds; left panel) and older infants (15- to 17-month-olds; right panel). Proportion of fixations to the older images when they were the target are indicated by the yellow lines, and fixations to the younger images when they were the target are indicated by the green lines. The critical window of analysis is marked by the light gray dotted lines, and the onset of the target noun is indicated by the black dotted line. Data is aggregated across participants and trials. Error bars represent standard error.

To further explore the effect of image type, we assessed the extent to which our pre-assigned image types were reported to be experienced more frequently by the intended age group. That is, we used our measure of caregiver-reported experience as a manipulation check of our image types. For each subject, we averaged experience across each condition to compute a measure of each infant's average experience with images relevant for older infants and their average experiences with images relevant for younger infants. We then fit an LMEM regressing this average experience on image type, age group, and their interaction, including a by-subject random intercept. If our pre-assignments of images into older and younger categories were correct, we would expect image type and age to interact such that images relevant for older infants are reported as more highly experienced by older infants and images relevant for younger infants are reported as more highly experienced by younger infants. However, there was no significant effect of image type, age, or their interaction (p=0.502, p=0.164, p=0.088, respectively), suggesting that our preassigned conditions were not representative of each age group's actual referent experience.

Given that our preassigned image types were not reflective of experience, we next fit a LMEM using caregiver-reported experience in the past week as the measure of referent experience. We regressed baseline corrected accuracy on experience (adaptively centered; Raudenbush, 2009), age group (coded as -0.5 for younger infants and 0.5 for older infants), and their interaction, including by-subject and by-item random intercepts. We used baseline-corrected accuracy in this model, as our prior results indicated baseline preferences (e.g., Figure 7). This model diverges from the preregistered model, an LMEM predicting accuracy from experience with age group only included as a covariate. We preregistered this set of analyses because we assumed that the first model (e.g., using image type as the experience measure) would test for experience and age interactions, while the second model would elucidate whether individual referent experience explained additional variance beyond the effect of age. Given the lack of correlation between our conditions and experience, this exploratory LMEM including an interaction term is a stronger test of our primary hypothesis. Furthermore, we can still assess the effect of experience alone by examining the simple effects in the current model.

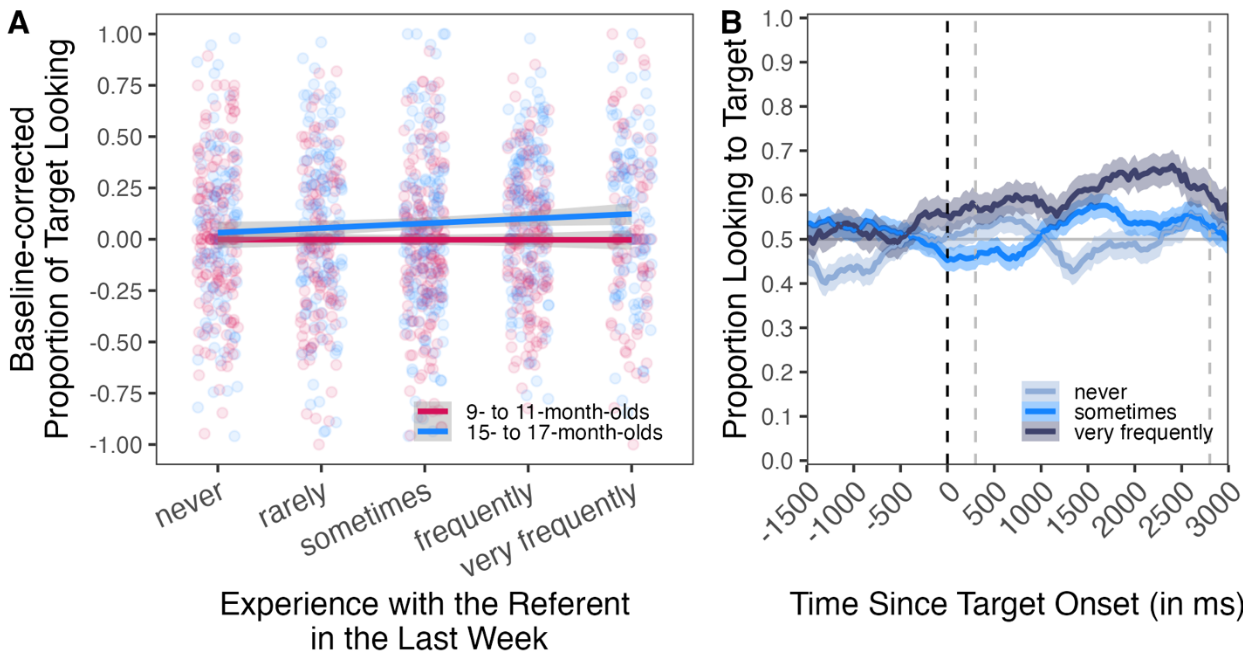
The LMEM model including caregiver-reported experience, age group, and their interaction, revealed a significant effect of age group (b=0.080, $F(1,72.39) = 16.967$, $p<0.05$) and a significant interaction (b=0.039, $F(1,909.54)=4.661$, p=0.031). Older infants were significantly more accurate at recognizing words than younger infants, especially for referents with which they had more experience (Figure 8A). The intercept was also significantly above zero

(M=0.034; b=0.037, $F(1,4.63)$=14.816, p=0.014), suggesting that, on average, infants increased their looking to the target image after hearing the noun relative to their looks to the target image before hearing the noun.

To further investigate the results, we examined the simple effects by recoding the age group variable so that each age bin served as the reference group (e.g., younger infants were recoded as 0 for one model and older infants were recoded as 0 for a separate model). Recoding the age group in this way allows us to test whether the effects hold for each age group. We found that the younger infants did not recognize words significantly above chance (M=-0.002; p=0.857). The older infants, however, did recognize the words on average (M=0.078; $p<0.05$). Critically, caregiver-reported experience with the referents was a significant predictor of accuracy for the older infants (Figure 8B; b=0.029, $F(1,1241.70)$=4.657, p=0.031), but not for the younger infants (p=0.147). These results suggest that older infants were better at recognizing words in the presence of referents that they were reported to experience often (Figure 8B).

Figure 8

The effect of recent experience on word recognition



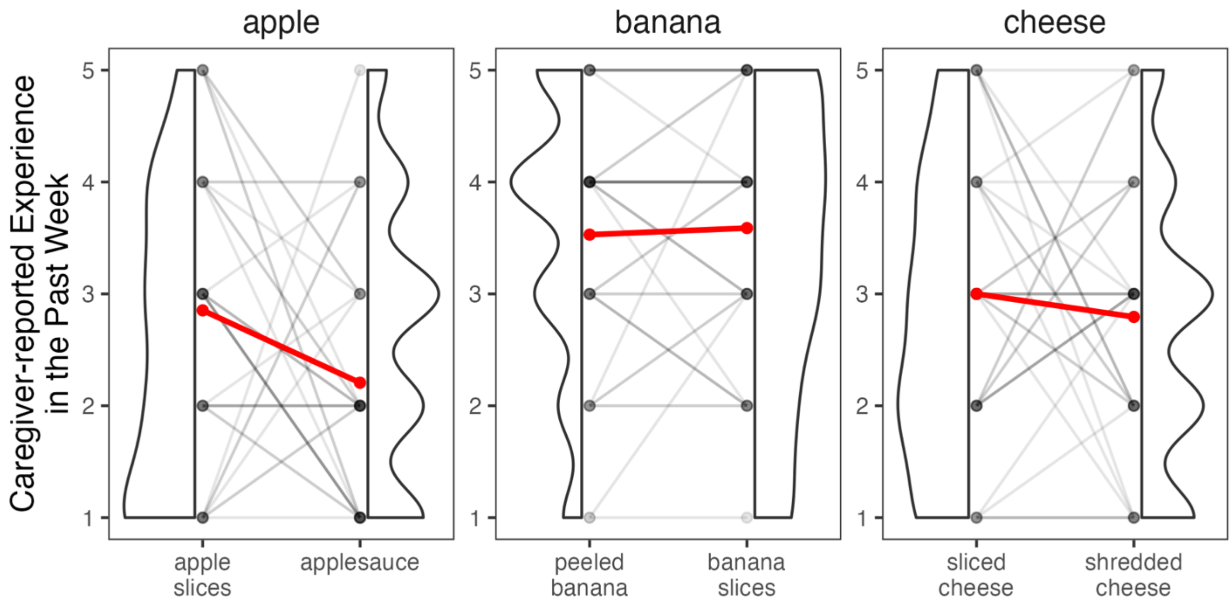
Note. The interaction effect of experience and age (2a) and the effect of experience on older infants (2b). Each data point in (2a) represents one trial of the LWL task, and (2b) shows data averaged across trials and participants for the older group only. Error bars represent standard error of the mean. The black dotted line in 2b indicates the onset of the noun, and the light gray dotted lines demarcate the critical window. In panel B, the proportion of fixation to the target referent is plotted for three different levels of experience: never experienced (light blue line), sometimes experienced (medium blue line), and very frequently experienced (dark blue line).

In addition to examining the role of experience on word recognition, we were interested in whether the data could provide initial evidence for individual differences in word meanings depending on infants' experiences with referents. There are two possible explanations for the experience-related effects we observed with the older infants. One is that differences in experience with specific referents drive individual differences in early word meanings. The other is that while experience does indeed affect early word meanings, infants have such similar experiences that those word meanings do not differ. We reasoned that the distribution of responses on the experience survey would help us to determine whether the present data provide potential evidence for individual differences in early word meanings, or whether, instead, infants' experiences overlap so much that experience-driven word meanings are likely to be highly similar. We examined this question by graphically representing caregiver reports of individual infants' experience with the two tested referents for each of the three words (Figure 9). If the effect of experience on infants' word recognition is due to shared experiences with referents (or lack thereof), we would expect some referents to be very frequently experienced by all participants, and others to be rated as rarely experienced. However, if infants' idiosyncratic experiences result in individual differences in word recognition, we would expect substantial variability in experience with each word's referents.

Visual inspection of Figure 9 suggests that infants' experiences with these referents are highly variable. The distributions of reported experience vary across the entire experience rating scale, and the data do not appear to be skewed towards a specific rating for any of the referents. Many infants show different levels of experience with the two tested referents for each word, as evidenced by steep slopes for individual infants. "Banana" shows the least variability; the two referents for the word "banana" are both experienced relatively frequently (peeled banana: M=3.529, SD=1.161; sliced banana: M=3.588, SD=1.158). By contrast, referents for the word "apple" are relatively infrequently experienced on average (apple slices: M=2.853, SD=1.351; applesauce: M=2.206, SD=1.149). The two referents for the word "cheese" are also experienced less frequently in general (sliced cheese: M=3.000, SD=1.326; shredded cheese: M=2.794, SD=1.250). Of particular relevance, the referents for the words "apple" and "cheese" seem to be experienced with a high degree of variability across participants. This observation is consistent with the hypothesis that the effects of experience on word recognition observed in this study are due to individual differences in word meaning representations.

Figure 9

Experience with referents for each word



Note. The graph depicts experience ratings for each of the referents. Each panel represents a specific word, and the gray lines connect experience ratings between the two referents for a given participant. The average ratings for each word are denoted by the red lines. The distribution of the data is represented by the half violin plot.

Discussion

Infants' experiences with objects are highly idiosyncratic. Food words, in particular, are likely to be experienced in various forms (e.g., puree, sliced, or whole) that differ across families and development. We aimed to harness this natural variability to understand whether and how prior experience shapes early word meanings. We measured 9- to 11-month-olds' and 15- to 17-month-olds' word comprehension in the presence of two different referents–a younger appropriate form and an older appropriate form–for three early learned food words. We also asked their caregiver to report the frequency with which their child experienced each form of food in the last week. Older infants recognized words significantly more accurately than younger infants, and this effect was even stronger for referents that they had frequently encountered. The younger infants demonstrated no evidence of word recognition regardless of experience. The older infants, by contrast, recognized words more accurately when the referent was reported to be frequently experienced in the last week. Put another way, older infants were better at linking recently and often experienced referents with familiar labels.

These findings suggest that word meanings are influenced by prior experience, which is similar to the role that experience plays in shaping adults' word meanings (Rodd et al., 2016; Wiley et al., 2018). For adults, recent experience shapes how strongly a given referent is associated with a label, allowing some meanings to be more readily retrieved than others. Infants' recent experience may also facilitate word recognition in the presence of specific referents by influencing word-referent associations. Infants have many opportunities to learn food words during mealtimes (Clerkin & Smith, 2022; Custode & Tamis-LeMonda, 2020; Tamis-Lemonda et al., 2018) and experience food in some forms more often than others (e.g., sliced apples rather than whole apples; Clerkin & Smith, 2022). Thus, the more frequently-encountered forms may be more strongly associated with the label than less frequently-encountered forms, ultimately, facilitating word comprehension for highly experienced referents in the current LWL task.

Our findings add to the rich literature on variable exemplars and word learning, demonstrating that the relative amount of exposure to a given referent influences the representations that infants develop. Infants experience individual instances of a category, and these referent-label associations are stored (Ambridge, 2020). As infants amass more variable referent experience, they learn to abstract and generalize known labels to novel referents (Crespo et al., 2024; Kucker & Samuelson, 2012; Perry et al., 2010). Furthermore, word learning unfolds gradually over time as infants slowly retain associations between words and referents (Hendrickson et al., 2017; Samuelson & McMurray, 2017). The current results take this theory one step further; as infants amass experience with words and their referents, some are more strongly associated with these learned representations than others.

We also found limited word recognition abilities in 9- to 11-month-olds. While there is mixed evidence that 9-month-olds successfully comprehend early food words (Bergelson & Swingley, 2012, 2015; Foushee & Srinivasan, 2024; Kartushina & Mayor, 2019; Steil et al., 2021), the present results are likely due to the relatively challenging nature of our LWL task. We paired images exhaustively within a single superordinate category (e.g., foods). For example, the word "apple" was tested in the presence of apple slices paired with a peeled banana or shredded cheese. Prior work, however, paired food words with referents from entirely separate superordinate categories (e.g., pairing cow and squash; Foushee & Srinivasan, 2024). These within-category comparisons may have been particularly challenging for our younger participants, as their word meanings are imprecise, often including related but incorrect referents (e.g., including sock for the word "foot"; Bergelson & Aslin, 2017). Thus, the younger age group likely did not have strong enough word representations to distinguish between referents within the same broader food category.

While the challenging task may have hindered younger infants' word recognition, it may also have allowed us to demonstrate experience effects in older infants. Prior research examining the role of experience has demonstrated that 14- to 18-month-olds have strong word recognition abilities, and their prior experience plays a limited role in word comprehension (Garrison et al., 2020; Weaver et al., 2024). By contrast, we found evidence that word recognition was better for more frequently experienced referents, perhaps because infants were discriminating between more challenging referent pairs. Put another way, identifying the target referent for the word "dog" in the presence of a golden retriever and a sparrow may be too simple a task to highlight subtle differences in word-referent associations. However, identifying the word "apple" in the presence of apple slices and banana slices (both of which are beige and cut-up pieces of fruit) may be a methodologically stronger test of experience-based improvements in word comprehension. Indeed, prior research demonstrates evidence for this methodological "desirable difficulty". For example, Southgate and Meints (2000) found evidence for a typicality effect when the pairs of images tested were within basic-level categories (e.g., a trial testing the word "bird" included a sparrow and an ostrich), but Weaver and colleagues (2024) found no evidence of typicality effects in cross-category test trials. Taken together, these results suggest that subtle individual differences may be more apparent when testing word comprehension under more challenging LWL tasks.

Limitations

Although the results provide evidence that prior experience influences 15- to 17-month-olds' word comprehension, our results do not necessarily indicate individual differences in early word meanings. The exploratory analysis examining the experience survey demonstrated immense variability in the frequency with which infants experience the different referents for the words "apple" and "cheese". Therefore, the effect of experience seen in the present study may be driven by individual differences in word meanings. Put another way, our results cannot simply be explained by some referents being generally more experienced by all of our infants, at least for the words "apple" and "cheese". However, most of the infants in the present study did, in fact, frequently encounter both referents for the word "banana". Thus, the present findings do not provide robust evidence of individual differences in early word meanings. Instead, our results are consistent with the hypothesis that infants associate some referents more strongly with known labels. Future work could address the question of individual differences in word meanings more directly by exposing infants to novel word-referent associations with varying frequency between subjects, and testing whether infants' word comprehension is superior in the presence of frequently-encountered referents.

General Discussion

Words refer to entire categories of referents, some of which may be more frequently experienced than others. The current research investigated the range of referent experiences for a set of early-learned nouns and whether these experience-based individual differences influence infants' early word meanings.

In Study 1, we developed a database of images depicting the referents that U.S. infants encounter. Caregivers provided insight into their child's object experiences by providing photographs for the database, and norming the resultant image set. There was remarkable variation in the number and kinds of referents provided for a given word. Despite this variation, pairs of referents provided for a given word were generally considered more similar than pairs of referents from different words, suggesting that categories of referents have overlapping perceptual features. We also found that referents were reported to be experienced occasionally on average, yet individual experiences with referents ranged the entire scale (i.e., from never to very frequently). These finding suggests that the final database is fairly representative of referent experiences during infancy in the United States. Of particular relevance, these findings demonstrate immense variability in both the specific referent that infants experience and in the amount of experience a given child has with referents.

In Study 2, we examined whether infants' word meanings were related to their experience with individual referents. To do so, we measured infants' comprehension of food words using a looking-while-listening paradigm. We intentionally included target referents that depicted the foods in two distinct forms (e.g., applesauce and apple slices), and asked caregivers to report how often their infant experienced each form in the past week. Younger infants demonstrated very little evidence of word recognition regardless of their experience with referents. Older infants, by contrast, were more accurate at recognizing words when the target image depicted a frequently-experienced referent. These findings reveal that infants associate some referents more strongly with their labels, depending on how frequently they are encountered.

The results indicate that the referent space and early word representations are potentially more variable than traditional word learning theories assume. That is, word learning has been framed as a process by which infants link a word to a referent in "buzzing, blooming confusion" (James, 1890) of possible different referents in the world. As an example, consider the Gavagai problem (Quine, 1960). Quine (1960) described word learning as akin to the challenge that a linguist faces when determining whether an exclamation ("Gavagai!") in an unknown language refers to a rabbit, an action, a phrase, or even a part of a rabbit. Notably absent from this example is whether "gavagai" refers to an angora rabbit and/or a cottontail rabbit. The present findings suggest that no single referent is necessary or sufficient to learn the meaning of a word. Instead, this research is consistent with the hypothesis that words have multiple related meanings that vary across individuals (Elman, 2009; Rosch & Mervis, 1975; Srinivasan & Rabagliati, 2021)–even at the early stages of word learning. That is, infants were reported to experience a multitude of different referents for a given word, some of which were more frequently encountered than others. Moreover, this variability in experience influenced infants' word comprehension.

Although these data do not provide direct evidence for individual differences in infants' word meanings, they do establish the necessary components for such a theory (e.g., variability in referents and asymmetric word-referent associations).

The current findings are consistent with two different hypotheses about the nature of early referent experience. On the one hand, infants may have very narrow referent experiences. By this account, infants only experience a select few referents for a word (e.g., one or two), but encounter those referents very frequently. Alternatively, referent experience may be broad, but asymmetric. Under this hypothesis, infants experience the entire category of referents, but the relative frequency of exposure to particular referents varies. Take, as an example, the word "dog". In Study 1, the most highly experienced referent for the word "dog" was a golden retriever. This high rating could be explained by infants only experiencing golden retrievers (narrow referent experience) or by most infants experiencing a variety of dog exemplars but encountering golden retrievers most often (broad referent experience). Our data cannot adjudicate between these two alternative explanations for early referent experience. This is an important distinction to elucidate in future work to understand precisely how word-referent experiences develop.

Broader impacts and future directions

The current results have several important methodological and theoretical implications for familiar word learning. In what follows, we focus on three primary avenues for future research. We discuss diverse language trajectories, moving beyond the "map trap", and using looking-while-listening studies to examine individual differences in light of the current findings.

Diverse language trajectories

Late talkers (LTs) are often defined as toddlers who have smaller productive vocabularies relative to their age-matched peers, despite otherwise typical cognitive development. An estimated 50% of LTs ultimately "catch up" to their typically-developing peers and do not have long-term challenges in word learning (Rescorla, 2002, 2011). A central question in this area is how to best measure early language and identify toddlers at risk for later language disorders (e.g., Developmental Language Disorder). One prevailing hypothesis is that toddlers with persistent language delays have vocabularies with different semantic structures than typically-developing toddlers (Colunga & Sims, 2017; Horvath et al., 2022; Jiménez & Hills, 2022; Perry et al., 2022). In particular, LTs' vocabularies seem to include fewer shape-based words (Perry et al., 2022), and they struggle to extend labels to novel objects that share a similar shape (Zuniga-Montanez & Krott, 2025).

While prior work suggests that the structure of LTs' vocabularies is different from typical-language toddlers, it leaves open the question of whether LTs' word meanings also vary from typically developing infants. LTs' challenges with the shape-bias suggest that they may struggle to associate or extend familiar labels to a broader category of referents. By this theory, we would expect LTs' word meanings to be even more tightly bound to the specific referents that they experience than typical-developing infants' word meanings. Interestingly, prior work indicates that LTs have slower and less accurate word recognition than their language-typical peers (Fernald & Marchman, 2012; LaTourrette et al., 2023). One possible explanation for these findings is that late talkers find it more difficult to identify a never-before-seen referent for a given word because of their difficulty extending labels to shape-matched exemplars. Alternatively, poorer word recognition abilities may be explained by broader challenges with associating words and referents. Future work could disentangle these competing hypotheses by examining whether LTs are faster and more accurate at recognizing a frequently-experienced referent than a novel, prototypical referent. If LTs' word recognition mirrors typically-developing toddlers' word recognition in the presence of frequently-encountered referents, we would interpret this as evidence that word extension may be a key indicator of later language delay. However, if late talkers do not have equally inferior word recognition for both referent types, we may interpret this as evidence that associating words and referents may be a challenge for LTs.

Moving beyond the "map trap"

Word learning theories have been limited by their conceptualization of familiar word learning as a mapping problem (Wojcik et al., 2022). Our studies emphasize the importance of moving beyond simple mappings between a word and a single referent to characterize the complexity of early word representations. A challenge for future research will be to understand which kinds of experiences matter for individual differences and whether individual differences are persistent across word classes.

As a first foray into experience-driven effects on early word-referent associations, we selected the frequency with which infants encountered different referents (i.e., visual experience) as our primary measure of experience across the two different studies. Defining experience in this specific manner had several advantages. First, recent reviews of familiar word learning have criticized the use of a researcher-intuited referent to represent a word's meaning (Wojcik et al., 2022) without providing alternative visual stimuli to address this issue. Our studies provide multiple visual referents for researchers, as well as concrete evidence that the specific images included in looking-while-listening studies impact word comprehension. Secondly, visual experience is a tangible, concrete, and straightforwardly observable property for caregivers to report. This contrasts with other ways of defining experience (e.g., familiarity), which may be interpreted in many ways (e.g., familiarity may conflate frequency and typicality) and have resulted in very limited effects of experience in prior work (Garrison et al., 2020). Finally, prior literature established high variation in the visual experiences in early infancy using naturalistic head-camera data (Clerkin et al., 2017; Clerkin & Smith, 2022). While limiting experience to the visual domain provided critical insights into variation among word-referent associations, there are several alternative ways to define experience.

Future research could consider whether variability in the contextual experiences under which infants learn familiar words impacts word meanings. Prior work demonstrates that variability in the location in which words are heard influences the ease with which a word is learned (Breitfeld & Saffran, 2025; Custode & Tamis-LeMonda, 2020; Roy et al., 2015; Tamis-Lemonda et al., 2018). Perhaps the location where a word is encountered influences how strongly infants associate that word with a given meaning. For instance, the word "chicken" can refer to both an animal and a food. Some infants may hear the word "chicken" more often during mealtimes, and therefore, primarily in the context of a highchair in the kitchen. Other infants may hear the word "chicken" more frequently during play, and thus, encounter the word "chicken" in the living room, bedroom, or playroom. If our experience findings generalize to other kinds of experience, we would predict that infants with more constrained contextual experiences would more strongly associate the word "chicken" with the sense used in that context (i.e., a food sense in the kitchen and an animal sense in other rooms). We would also predict that the proportion of these contexts would matter for word representations, such that more widespread exposure across different contexts would result in a broader word meaning for that child.

Another fruitful direction to explore is whether differences in language experience influence word meanings. Specifically, multilingual or multi-dialectal input may result in more variable word meanings than monolingual input. This may be especially true for infants who experience input from multiple languages in separate contexts. For instance, a child who is primarily exposed to Spanish speakers at home, but is exposed to English speakers at daycare, may experience distinctive referents in each of these contexts (e.g., their own dog at home, but a stuffed animal dog at daycare). This experience may result in stronger associations between a word in a specific language and the referent that is most commonly experienced in that language context. A potential method to investigate this question would be to test multilingual infants' word comprehension in both languages for a set of concepts that are experienced in both contexts (e.g., dog). The referents included in the task could be drawn directly from each language context (e.g., the pet dog from home and the stuffed daycare dog). Infants' speed and accuracy of word recognition for each referent in the presence of each language would be measured. To use the previous example, we would examine whether an infant was equally fast and accurate at recognizing the word "perro" in the presence of the stuffed dog from daycare as they are at recognizing it when hearing the word "dog". Given the current findings, we would predict stronger associations between referents and the specific language in which they experience the referent.

In addition to considering different kinds of experience, investigating word meanings for abstract words rather than concrete, imageable nouns will advance theories of familiar word learning. The literature on familiar words has traditionally focused on nouns, characterizing language development by the size and structure of noun vocabularies. However, nouns do not dominate the early vocabularies of all languages equally (e.g., Mandarin; Tardif et al., 1997). Thus, one possible way to include more diverse populations in early language development research is to diversify the kinds of words tested.

Words that are considered more abstract than concrete nouns (e.g., abstract nouns, verbs, and adjectives) have been theorized to be learned later, and can be more challenging to assess using familiar word paradigms (Bergelson & Swingley, 2013, 2015; Casey et al., 2023; Nomikou et al., 2019). One of the persistent challenges in testing abstract words is that the referents are harder to identify and represent pictorially. For example, Bergelson and Swingley (2013, 2015) investigated whether very young infants would recognize verb meanings when using video stimuli. While word recognition for nouns was similar across static images and dynamic videos, infants did not recognize verbs in the presence of videos with a target action. By contrast, Nomikou and colleagues (2019) used static images to test verbs and found a significant increase in target looking after hearing a verb compared to before hearing the verb (although they did not compare accuracy to chance). These examples demonstrate that the specific stimuli included for abstract words may impact the results, yet there are very limited recommendations on how to select stimuli when testing abstract words. However, we have established a novel method for collecting images by asking caregivers to provide photographs that portray their infants' experiences with referents. While concrete nouns were relatively easy to photograph, it is an open question whether we could leverage caregiver report methods to collect stimuli for the entire set of words on the MB-CDI words, including verbs, adjectives, and other abstract words. One challenge of the method in Study 1 for abstract words is that abstract words may be more likely to include identifiable information (e.g., a child jumping as a referent for jump). Future work could potentially circumvent this challenge by collecting referent images from families and subsequently recreating the images without identifiable information.

Despite the challenges in developing adequate methods for testing abstract words, they may be more likely to have variable word meanings. Indeed, prior research examining adults' word meanings suggests that abstract words vary more across individuals than concrete nouns (Wang & Bi, 2021). This suggests that abstract words may provide important insights into individual variation in early word meanings. Therefore, an important next step in examining early word meanings would be to identify candidate abstract words to examine the role of experience-based individual differences in infants' word-referent associations. A potential starting point may be to use a data-driven approach by which we examine the range of referents provided by caregivers and select abstract words that appear to have a wide range of reported experience and/or low pairwise similarity within a category.

Methodological insights

The findings from Study 1 suggest that several different referents can be included as targets in LWL tasks. Importantly, future studies should include referents according to their central hypotheses rather than relying on a rule-based approach to referent selection (e.g., always use the highest average rated image). For example, if a researcher aims to demonstrate evidence of familiar word recognition at particular age ranges, the present results indicate that using a highly experienced referent may be the strongest test of this ability. Researchers could include either a referent with a high modal experience rating or a referent with the highest mean experience rating. Alternatively, a study aiming to examine the richness of representations for particular words should include multiple referents across the range of experiences. This approach would allow the researcher to investigate whether infants vary in how they associate referents across a range of experiences. More broadly, the database provides norms that can be used to select visual stimuli in a theory-driven manner rather than relying on researchers' intuitions.

The findings also provide methodological insights for investigating individual differences in word recognition. Individual differences in early language development are notoriously challenging to investigate (Donnelly & Kidd, 2020), with the majority of the familiar word literature examining group differences. For instance, several studies have identified differences in semantic structure between the vocabularies of infants with different language abilities (e.g., Perry et al., 2022). Study 2, by contrast, examined relations between individual infants' referent experiences for specific words. We found that infants encounter referents with different frequencies, which shapes their word-referent associations. We may have been able to detect individual differences due to three methodological decisions: (1) we included two discontinuous age bins; (2) we developed a challenging LWL task; (3) we included multiple measures of word recognition.

In Study 2, we examined two distinct, discontinuous age groups, each with an average sample size (N=~40). Using this design, we found differences in the effect of experience across the two age ranges. Prior studies, however, have used continuous age ranges (Garrison et al., 2020; Weaver et al., 2024). Overall increases in word recognition across a broad developmental time may mask experience effects for particular age groups. Moreover, studies using larger age ranges often dichotomize their continuous age variable into two age bins, typically varying in the number of participants in each group (Bergelson & Swingley, 2012; Garrison et al., 2020). Dichotomization of continuous variables reduces statistical power (Royston et al., 2006) and arbitrarily defines older and younger infants (e.g., 13.9 is younger but 14.0 is older). By selecting theoretically-motivated discontinuous age ranges, we were able to elucidate different effects of experience. Our findings suggest that cross-sectional designs can be useful in identifying developmental changes in experience effects, but that the study must be sufficiently powered to detect these effects. Additionally, the results indicate that future studies should consider recruiting relatively equal samples of infants in each age bin of interest.

Another explanation for detecting experience effects in Study 2 was the difficulty of the looking-while-listening task. Infants had to recognize words in the presence of two images from the same superordinate category, which is known to be more challenging (Arias-Trejo & Plunkett, 2010; Ellis Weismer et al., 2016). However, the relative difficulty of our experiment may have allowed infants to demonstrate the experience effect (which we predicted to be small). That is, experience may play a small role in word recognition when infants need to identify a word in the presence of two objects that have remarkably different perceptual features (e.g., recognizing “dog” in the presence of a dog and a bird). Under more challenging task designs, infants may need to draw on a more precise word representation to identify the target image. Given these findings, one important consideration for future studies is whether the study is sufficiently difficult to elucidate the effect of interest. Fortunately, the database developed in Study 1 will significantly improve researchers’ ability to scale the difficulty of their tasks by including pairwise similarity for images across and within word categories.

A final methodological consideration concerns the outcome measure used to index word meanings. We initially preregistered using the proportion of target looking during a critical window as our measure of word recognition. We had to modify this approach as infants showed baseline preferences, despite the referents appearing to be equally salient to the researcher (i.e., referents were all beige, from the same superordinate category, and of similar size). Yet, infants still preferred some stimuli over others. These data indicate that infants’ visual preferences may be challenging to predict *a priori*. As an alternative, it may be good practice to examine the robustness of the results across multiple pre-registered outcome measures. Both the current results and results from Weaver and colleagues (2024) suggest that the proportion of target looking and baseline-corrected proportion target looking are important measures to include. Weaver and colleagues (2024) and Weaver and Saffran (under review) have also shown that reaction time provides converging evidence for a given result. Unfortunately, we were unable to conduct analyses using reaction time in Study 2 because we did not have sufficient data after removing trials in which infants were initially fixating the target image. Future work should identify ways to modify LWL tasks to increase the number of trials included for reaction time studies, as well as include other alternative measures of looking behavior.

Conclusion

When hearing a word, infants may experience different referents. These experiences may vary across individual word learning moments and across children. The present studies represent a first step in characterizing this immense variability across a large set of early learned words. These results shed light on how idiosyncrasies in exposure to particular referents can manifest as differences in the strength of word-referent associations for individual infants. Understanding the extent to which word meanings vary for individual infants and words will help to broaden the scope of future word learning theories to better account for variation in early word representations.

Appendix A

Concrete imageable nouns from the MB-CDI

MB-CDI Category	words
Animal	bird, dog, duck, kitty, lion, mouse
Clothing	hat, pants, shoe, sock
Food and drink	apple, banana, bread, candy, cereal, cheese, cookie, juice, toast, water
Furniture and rooms	chair, couch, table, television (TV)
Outside and places	flower, moon, rock
Small household items	blanket, bottle, cup, dish, lamp, plant, radio, spoon
Toys	ball, book, doll
Vehicles	bar, stroller

Appendix B

Norms reported at the word category

Word	N	*H* value	Average Experience Rating
apple	26	1.178	3.307
ball	30	0.848	3.340
banana	23	1.559	3.400
bird	24	2.175	2.926
blanket	32	2.215	3.105
book	32	1.116	3.379
bottle	27	2.184	3.222
bread	27	1.052	3.136
candy	26	2.131	2.948
car	33	1.394	3.113
cereal	28	2.277	2.854
chair	32	1.192	3.168
cheese	28	1.140	3.076
cookie	22	1.726	3.157
couch	31	1.188	3.081
cup	32	1.476	3.137
dish	32	1.646	3.019
dog	30	2.090	3.221
doll	23	2.095	3.209
duck	25	2.173	3.116
flower	27	1.343	3.129
hat	30	1.186	3.174
juice	24	1.620	2.968
kitty	27	2.026	3.065
lamp	26	1.168	2.974
lion	22	1.985	2.930
moon	10	1.838	3.127
mouse	11	2.314	3.079
pants	31	1.341	3.411
plant	26	1.369	3.072
radio	11	2.064	2.835
rock	20	1.323	2.970
shoe	31	1.548	3.311
sock	30	1.487	3.421
spoon	31	0.647	3.474
stroller	20	1.499	3.275
table	30	0.575	3.184
television	31	2.083	3.151
toast	13	1.107	3.404
water	29	2.024	3.176

Appendix C

Sample Database entries for the words “apple”, “banana”, and “cheese”

Image	word	Image identifier	N	Modal name	*H* value	Exp. Rating (*M*)	Exp. Rating (Mo)	Exp. Rating (*SD*)
	apple	apple01	20	apple	0.000	3.000	1	1.897
	apple	apple02	30	apple	0.561	3.367	5	1.790
	apple	apple03	16	apple	0.669	4.500	5	1.150
	apple	apple04	25	juice	3.075	2.828	4	1.256
	apple	apple05	30	apple	0.833	3.900	5	1.729
	apple	apple06	28	apple	0.959	3.452	5	1.609
	apple	apple07	24	apple	1.143	3.042	1	1.853

	apple	apple08	30	apple	0.970	3.387	5	1.838
	apple	apple09	29	apple	1.280	3.710	5	1.553
	apple	apple10	30	apple	0.833	3.903	5	1.535
	apple	apple11	30	apple	1.766	2.677	1	1.492
	apple	apple12	31	apple	1.531	3.594	5	1.720
	apple	apple13	22	apple	0.000	3.720	5	1.768
	apple	apple14	29	apple	1.677	3.448	5	1.863
	apple	apple16	32	apple	0.597	2.727	1	1.773

	apple	apple17	30	apple	1.239	2.839	1	1.968
	apple	apple18	30	apple	2.663	2.531	2	1.270
	apple	apple19	31	apple	1.012	3.258	5	1.825
	apple	apple20	27	apple	0.455	3.889	5	1.577
	apple	apple21	22	apple	1.534	3.160	1	1.772
	apple	apple22	30	apple	1.442	3.233	5	1.832
	apple	apple23	30	apple	0.420	3.867	5	1.655
	apple	apple24	22	apple	0.267	3.318	5	1.783

	apple	apple25	26	pouch	3.536	2.571	2	1.476
	apple	apple26	16	apple	0.993	3.125	1	1.746
	banana	banana01	23	banana	1.492	3.296	5	1.772
	banana	banana02	27	banana	2.724	3.407	3	1.309
	banana	banana03	26	banana	0.850	3.692	5	1.850
	banana	banana04	29	banana	1.477	3.667	5	1.807
	banana	banana05	30	banana	0.675	3.290	5	1.987
	banana	banana06	31	banana	0.947	3.469	5	1.796

	banana	banana07	30	banana	2.878	2.677	2	1.249
	banana	banana08	19	banana	2.078	3.333	5	1.653
	banana	banana09	29	banana	1.563	2.645	1	1.561
	banana	banana10	27	banana	2.080	3.444	5	1.502
	banana	banana11	28	banana	1.370	3.786	5	1.813
	banana	banana12	26	banana	1.193	2.821	1	1.611
	banana	banana13	27	banana	1.247	4.074	5	1.685
	banana	banana14	30	banana	1.315	3.813	5	1.731

	banana	banana15	27	banana	0.725	3.517	5	1.883
	banana	banana16	28	banana	1.092	3.750	5	1.713
	banana	banana17	27	banana	1.975	2.750	3	1.295
	banana	banana19	29	banana	1.513	4.267	5	1.507
	banana	banana20	24	banana	1.356	3.652	5	1.722
	banana	banana21	29	banana	1.477	3.400	5	1.886
	banana	banana22	24	banana	3.137	2.080	2	1.038
	banana	banana23	28	banana	1.134	3.897	5	1.718

	cheese	cheese03	27	cheese	0.901	2.821	1	1.565
	cheese	cheese04	30	cheese	1.892	3.531	4	1.319
	cheese	cheese05	30	cheese	1.478	3.467	4	1.502
	cheese	cheese07	26	cheese	1.153	3.577	5	1.447
	cheese	cheese08	29	cheese	1.608	3.069	4	1.462
	cheese	cheese10	21	cheese	1.092	3.500	4	1.401
	cheese	cheese11	28	cheese	1.260	3.269	5	1.638
	cheese	cheese12	22	cheese	0.468	3.000	1	1.710

	cheese	cheese13	26	cheese	2.119	3.304	5	1.396
	cheese	cheese14	23	cheese	0.813	2.900	1	1.626
	cheese	cheese15	31	cheese	0.748	2.875	1	1.601
	cheese	cheese16	31	cheese	0.420	2.903	1	1.578
	cheese	cheese17	29	cheese	0.954	2.760	1	1.589
	cheese	cheese18	25	cheese	0.833	2.733	1	1.574
	cheese	cheese19	30	cheese	1.270	2.931	1	1.751
	cheese	cheese20	29	cheese	1.239	3.207	4	1.497

	cheese	cheese21	30	cheese	0.880	3.211	1	1.813
	cheese	cheese22	19	cheese	1.209	3.121	4	1.576
	cheese	cheese23	31	cheese	0.970	2.875	2	1.238
	cheese	cheese24	30	cheese	1.299	2.714	1	1.536
	cheese	cheese25	26	cheese	0.983	3.125	1	1.727
	cheese	cheese26	24	cheese	0.767	2.806	1	1.424

References

Ackermann, L., Hepach, R., & Mani, N. (2020). Children learn words easier when they are interested in the category to which the word belongs. *Developmental Science*, *23*(3), e12915. https://doi.org/10.1111/DESC.12915

Adlington, R. L., Laws, K. R., & Gale, T. M. (2009). The Hatfield Image Test (HIT): a new picture test and norms for experimental and clinical use. *Journal of Clinical and Experimental Neuropsychology*, *31*(6), 731–753. https://doi.org/10.1080/13803390802488103

Adobe Inc. (2025). *Adobe Photoshop (version 2025)*.

Ambridge, B. (2020). Against stored abstractions: A radical exemplar model of language acquisition. *First Language*, *40*(5–6), 509–559. https://doi.org/10.1177/0142723719869731/ASSET/IMAGES/LARGE/10.1177_0142723719869731-FIG2.JPEG

Arias-Trejo, N., & Plunkett, K. (2010). The effects of perceptual similarity and category membership on early word-referent identification. *Journal of Experimental Child Psychology*, *105*(1–2), 63–80. https://doi.org/10.1016/j.jecp.2009.10.002

Barry, C., Morrison, C. M., & Ellis, A. W. (1997). Naming the Snodgrass and Vanderwart Pictures: Effects of Age of Acquisition, Frequency, and Name Agreement. *The Quarterly Journal of Experimental Psychology: Section A*, *50*(3), 560–585. https://doi.org/10.1080/783663595

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/JSS.V067.I01

Bennetts, S. K., Mensah, F. K., Westrupp, E. M., Hackworth, N. J., & Reilly, S. (2016). The agreement between parent-reported and directly measured child language and parenting behaviors. *Frontiers in Psychology*, *7*(NOV), 1710. https://doi.org/10.3389/FPSYG.2016.01710/BIBTEX

Bergelson, E. (2020). The Comprehension Boost in Early Word Learning: Older Infants Are Better Learners. *Child Development Perspectives*, *14*(3), 142. https://doi.org/10.1111/CDEP.12373

Bergelson, E., Amatuni, A., Dailey, S., Koorathota, S., & Tor, S. (2019). Day by day, hour by hour: Naturalistic language input to infants. *Developmental Science*, *22*(1), e12715. https://doi.org/10.1111/DESC.12715

Bergelson, E., & Aslin, R. (2017). Semantic Specificity in One-Year-Olds' Word Comprehension. *Language Learning and Development*, *13*(4), 481–501. https://doi.org/10.1080/15475441.2017.1324308

Bergelson, E., & Swingley, D. (2012). At 6-9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(9), 3253–3258. https://doi.org/10.1073/PNAS.1113380109/SUPPL_FILE/PNAS.201113380SI.PDF

Bergelson, E., & Swingley, D. (2013). The acquisition of abstract words by young infants. *Cognition*, *127*(3), 391–397. https://doi.org/10.1016/J.COGNITION.2013.02.011

Bergelson, E., & Swingley, D. (2015). Early Word Comprehension in Infants: Replication and Extension. *Language Learning and Development*, *11*(4), 369–380. https://doi.org/10.1080/15475441.2014.979387

Berman, S., Friedman, D., Hamberger, M., & Snodgrass, J. G. (1989). Developmental picture norms: Relationships between name agreement, familiarity, and visual complexity for child and adult ratings of two sets of line drawings. *Behavior Research Methods, Instruments, & Computers*, *21*(3), 371–382. https://doi.org/10.3758/BF03202800/METRICS

Borovsky, A. (2022). Drivers of Lexical Processing and Implications for Early Learning. *Annual Review of Developmental Psychology*, *4*(1), 21–40. https://doi.org/10.1146/ANNUREV-DEVPSYCH-120920-042902

Borovsky, A., Ellis, E. M., Evans, J. L., & Elman, J. L. (2016a). Lexical leverage: category knowledge boosts real-time novel word recognition in 2-year-olds. *Developmental Science*, *19*(6), 918–932. https://doi.org/10.1111/DESC.12343

Borovsky, A., Ellis, E. M., Evans, J. L., & Elman, J. L. (2016b). Semantic Structure in Vocabulary Knowledge Interacts with Lexical and Sentence Processing in Infancy. *Child Development*, *87*(6), 1893–1908. https://doi.org/10.1111/CDEV.12554

Braginsky, M., Yurovsky, D., Marchman, V. A., & Frank, M. C. (2019). Consistency and variability in children’s word learning across languages. *Open Mind*, *3*, 52–67. https://doi.org/10.1162/OPMI_A_00026,

Brauer, M., & Curtin, J. J. (2018). Linear mixed-effects models and the analysis of nonindependent data: A unified framework to analyze categorical and continuous independent variables that vary within-subjects and/or within-items. *Psychological Methods*, *23*(3), 389–411. https://doi.org/10.1037/MET0000159

Breitfeld, E., & Saffran, J. R. (2024). Early word learning is influenced by physical environments. *Child Development*, *95*(3), 962–971. https://doi.org/10.1111/CDEV.14046

Breitfeld, E., & Saffran, J. R. (2025). Environmental context scaffolds children’s semantic representation of novel words. *Cognition*, *262*, 106162. https://doi.org/10.1016/J.COGNITION.2025.106162

Bricker, D., Squires, J., Mounts, L., LaWanda Potter, M., Robert Nickel, M., Twombly, E., & Farrell, J. (1999). Ages & Stages Questionnaire. *Baltimore, MD: Paul H. Brookes.* www.brookespublishing.com

Brodeur, M. B., Dionne-Dostie, E., Montreuil, T., & Lepage, M. (2010). The Bank of Standardized Stimuli (BOSS), a New Set of 480 Normative Photos of Objects to Be Used as Visual Stimuli in Cognitive Research. *PLOS ONE*, *5*(5), e10773. https://doi.org/10.1371/JOURNAL.PONE.0010773

Brodeur, M. B., Guérard, K., & Bouras, M. (2014). Bank of Standardized Stimuli (BOSS) Phase II: 930 New Normative Photos. *PLoS ONE*, *9*(9). https://doi.org/10.1371/JOURNAL.PONE.0106953

Campbell, J., & Hall, D. G. (2022). The scope of infants' early object word extensions. *Cognition*, *228*, 105210. https://doi.org/10.1016/J.COGNITION.2022.105210

Cannard, C., Blaye, A., Scheuner, N., & Bonthoux, F. (2005). Picture naming in 3- To 8-year-old French children: Methodological considerations for name agreement. *Behavior Research Methods*, *37*(3), 417–425. https://doi.org/10.3758/BF03192710/METRICS

Casey, K., Potter, C. E., Lew-Williams, C., & Wojcik, E. H. (2023). Moving Beyond "Nouns in the Lab": Using Naturalistic Data to Understand Why Infants' First Words Include Uh-Oh and Hi. *Developmental Psychology*, *59*(11), 2162–2173. https://doi.org/10.1037/DEV0001630

Chen, C.H., Houston, D. M., & Yu, C. (2021). Parent–Child Joint Behaviors in Novel Object Play Create High-Quality Data for Word Learning. *Child Development*, *92*(5), 1889–1905. https://doi.org/10.1111/CDEV.13620

Clerkin, E. M., Hart, E., Rehg, J. M., Yu, C., & Smith, L. B. (2017). Real-world visual statistics and infants' first-learned object names. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1711). https://doi.org/10.1098/RSTB.2016.0055

Clerkin, E. M., & Smith, L. B. (2022). Real-world statistics at two timescales and a mechanism for infant learning of object names. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(18), e2123239119. https://doi.org/10.1073/PNAS.2123239119/SUPPL_FILE/PNAS.2123239119.SAPP.PDF

Colunga, E., & Sims, C. E. (2017). Not Only Size Matters: Early-Talker and Late-Talker Vocabularies Support Different Word-Learning Biases in Babies and Networks. *Cognitive Science*, *41*(1), 73–95. https://doi.org/10.1111/COGS.12409

Crespo, K., Vlach, H., & Kaushanskaya, M. (2024). The effects of speaker and exemplar variability in children's cross-situational word learning. *Psychonomic Bulletin and Review*, 1–11. https://doi.org/10.3758/S13423-023-02444-6/TABLES/3

Custode, S. A., & Tamis-LeMonda, C. (2020). Cracking the code: Social and contextual cues to language input in the home environment. *Infancy*, *25*(6), 809–826. https://doi.org/10.1111/INFA.12361

D'Amico, S., Devescovi, A., & Bates, E. (2001). Picture Naming and Lexical Access in Italian Children and Adults. *Journal of Cognition and Development*, *2*(1), 71–105. https://doi.org/10.1207/S15327647JCD0201_4

Donnelly, S., & Kidd, E. (2020). Individual differences in lexical processing efficiency and vocabulary in toddlers: A longitudinal investigation. *Journal of Experimental Child Psychology*, *192*, 104781. https://doi.org/10.1016/J.JECP.2019.104781

Douglas, B. D., Ewell, P. J., & Brauer, M. (2023). Data quality in online human-subjects research: Comparisons between MTurk, Prolific, CloudResearch, Qualtrics, and SONA. *PloS One*, *18*(3). https://doi.org/10.1371/JOURNAL.PONE.0279720

Dunn, L. M., & Dunn, D. M. (2007). *Peabody Picture Vocabulary Test* (4th ed.). Pearson. https://www.pearsonassessments.com/store/usassessments/en/Store/Professional-Assessments/Academic-Learning/Brief/Peabody-Picture-Vocabulary-Test-%7C-Fourth-Edition/p/100000501.html

Ellis Weismer, S., Haebig, E., Edwards, J., Saffran, J., & Venker, C. E. (2016). Lexical Processing in Toddlers with ASD: Does Weak Central Coherence Play a Role? *Journal of Autism and Developmental Disorders*, *46*(12), 3755–3769. https://doi.org/10.1007/s10803-016-2926-y

Elman, J. L. (2009). On the Meaning of Words and Dinosaur Bones: Lexical Knowledge Without a Lexicon. *Cognitive Science*, *33*(4), 547–582. https://doi.org/10.1111/J.1551-6709.2009.01023.X

Enders, C. K., & Tofighi, D. (2007). Centering Predictor Variables in Cross-Sectional Multilevel Models: A New Look at an Old Issue. *Psychological Methods*, *12*(2), 121–138. https://doi.org/10.1037/1082-989X.12.2.121

Fausey, C. M. (2022). Sampling Everyday Infancy: Lessons and Questions. *Journal of Vision*, *22*(14), 3202–3202. https://doi.org/10.1167/JOV.22.14.3202

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., Tomasello, M., Mervis, C. B., & Stiles, J. (1994). Variability in Early Communicative Development. *Monographs of the Society for Research in Child Development*, *59*(5), i-185. https://doi.org/10.2307/1166093

Fenson, L., Marchman, V. A., Donna J Thal, P. D., Phillip S Dale, P. D., J Steven Reznick, P. D., & Elizabeth Bates, P. D. (2007). *MacArthur-Bates Communicative Development Inventories*. Paul H. Brookes Publishing Co.

Fernald, A., & Marchman, V. A. (2012). Individual Differences in Lexical Processing at 18 Months Predict Vocabulary Growth in Typically Developing and Late-Talking Toddlers. *Child Development*, *83*(1), 203–222. https://doi.org/10.1111/J.1467-8624.2011.01692.X

Fernald, A., Perfors, A., & Marchman, V. A. (2006). Picking Up Speed in Understanding: Speech Processing Efficiency and Vocabulary Growth Across the 2nd Year. *Developmental Psychology*, *42*(1), 98. https://doi.org/10.1037/0012-1649.42.1.98

Fernald, A., Pinto, J. P., Swingley, D., Weinberg, A., & McRoberts, G. W. (1998). Rapid Gains in Speed of Verbal Processing by Infants in the 2nd Year. *Psychological science, 9*(3), 228–231. https://doi.org/10.1111/1467-9280.00044

Fernald, A., Zangl, R., Portillo, A. L., & Marchman, V. A. (2008). Looking while listening: Using eye movements to monitor spoken language. *Developmental Psycholinguistics: On-Line Methods in Children's Language Processing*, 97–135. https://www.researchgate.net/publication/267179424

Foushee, R., & Srinivasan, M. (2024). Infants who are rarely spoken to nevertheless understand many words. *Proceedings of the National Academy of Sciences of the United States of America*, *121*(23), e2311425121. https://doi.org/10.1073/PNAS.2311425121/SUPPL_FILE/PNAS.2311425121.SAPP.PDF

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2017). Wordbank: an open repository for developmental vocabulary data. *Journal of Child Language*, *44*(3), 677–694. https://doi.org/10.1017/S0305000916000209

Fu, S., Tamir, N., Sudaram, S., Chai, L., Dekel, T., & Isola, P. (2023). DreamSim: Learning New Dimensions of Human Visual Similarity using Synthetic Data. *arXiv Preprint*.

Ganea, P. A., Allen, M. L., Butler, L., Carey, S., & DeLoache, J. S. (2009). Toddlers' referential understanding of pictures. *Journal of Experimental Child Psychology*, *104*(3), 283–295. https://doi.org/10.1016/J.JECP.2009.05.008

Garrison, H., Baudet, G., Breitfeld, E., Aberman, A., & Bergelson, E. (2020). Familiarity plays a small role in noun comprehension at 12–18 months. *Infancy*, *25*(4), 458–477. https://doi.org/10.1111/INFA.12333

Golinkoff, R. M., Ma, W., Song, L., & Hirsh-Pasek, K. (2013). Twenty-Five Years Using the Intermodal Preferential Looking Paradigm to Study Language Acquisition: What Have We Learned? *Perspectives on Psychological Science*, *8*(3), 316–339. https://doi.org/10.1177/1745691613484936

Google. (2024). *Gemini AI*.

Hebart, M. N., Dickter, A. H., Kidder, A., Kwok, W. Y., Corriveau, A., Van Wicklin, C., & Baker, C. I. (2019). THINGS: A database of 1,854 object concepts and more than 26,000 naturalistic object images. *PLOS ONE*, *14*(10), e0223792. https://doi.org/10.1371/JOURNAL.PONE.0223792

Hendrickson, K., Poulin-Dubois, D., Zesiger, P., & Friend, M. (2017). Assessing a continuum of lexical–semantic knowledge in the second year of life: A multimodal approach. *Journal of Experimental Child Psychology*, *158*, 95–111. https://doi.org/10.1016/J.JECP.2017.01.003

Hintz, F., Meyer, A. S., & Huettig, F. (2017). Predictors of verb-mediated anticipatory eye movements in the visual world. *Journal of Experimental Psychology: Learning Memory and Cognition*, *43*(9), 1352–1374. https://doi.org/10.1037/XLM0000388

Hirsh-Pasek, K., Cauley, K. M., Golinkoff, R. M., & Gordon, L. (1987). The eyes have it: lexical and syntactic comprehension in a new paradigm. *Journal of Child Language*, *14*(1), 23–45. https://doi.org/10.1017/S030500090001271X

Horst, J. S., & Hout, M. C. (2016). The Novel Object and Unusual Name (NOUN) Database: A collection of novel images for use in experimental research. *Behavior Research Methods*, *48*(4), 1393–1409. https://doi.org/10.3758/S13428-015-0647-3/FIGURES/9

Horvath, S., Kueser, J. B., Kelly, J., & Borovsky, A. (2022). Difference or Delay? Syntax, Semantics, and Verb Vocabulary Development in Typically Developing and Late-talking Toddlers. *Language Learning and Development*, *18*(3), 352–376. https://doi.org/10.1080/15475441.2021.1977645

Houston-Price, C., Mather, E., & Sakkalou, E. (2007). Discrepancy between parental reports of infants' receptive vocabulary and infants' behaviour in a preferential looking task. *Journal of Child Language*, *34*(4), 701–724. https://doi.org/10.1017/S0305000907008124

Huettig, F., & Altmann, G. T. M. (2005). Word meaning and the control of eye fixation: semantic competitor effects and the visual world paradigm. *Cognition*, *96*(1), B23–B32. https://doi.org/10.1016/J.COGNITION.2004.10.003

Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, *137*(2), 151–171. https://doi.org/10.1016/J.ACTPSY.2010.11.003

Jago, L. S., Alcock, K., Meints, K., Pine, J. M., & Rowland, C. F. (2023). Language outcomes from the UK-CDI Project: can risk factors, vocabulary skills and gesture scores in infancy predict later language disorders or concern for language development? *Frontiers in Psychology*, *14*, 1167810. https://doi.org/10.3389/FPSYG.2023.1167810/BIBTEX

James, W. (1890). The principles of psychology, Vol I. In *The principles of psychology*. Henry Holt and Co. https://doi.org/10.1037/10538-000

Jiménez, E., & Hills, T. T. (2022). Differences in the Semantic Structure of the Speech Experienced by Late Talkers, Late Bloomers, and Typical Talkers. *Developmental Psychology*, *59*(1), 141–160. https://doi.org/10.1037/DEV0001446

Karmel, B. Z. (1969). The effect of age, complexity, and amount of contour on pattern preferences in human infants. *Journal of Experimental Child Psychology*, *7*(2), 339–354. https://doi.org/10.1016/0022-0965(69)90055-1

Kartushina, N., & Mayor, J. (2019). Word knowledge in six- to nine-month-old Norwegian infants? Not without additional frequency cues. *Royal Society Open Science*, *6*(9). https://doi.org/10.1098/RSOS.180711

Knabe, M. L., Schonberg, C. C., & Vlach, H. A. (2023). Does the public know what researchers know? Perceived task difficulty impacts adults' intuitions about children's early word learning. *Cognitive Research: Principles and Implications*, *8*(1), 1–18. https://doi.org/10.1186/S41235-023-00493-Y/TABLES/2

Kosie, J. E., & Lew-Williams, C. (2024). Infant-directed communication: Examining the many dimensions of everyday caregiver-infant interactions. *Developmental Science*, e13515. https://doi.org/10.1111/DESC.13515

Krautz, A. E., & Keuleers, E. (2022). LinguaPix database: A megastudy of picture-naming norms. *Behavior Research Methods*, *54*, 941–954. https://doi.org/10.3758/s13428-021-01651-0/Published

Kucker, S. C., & Samuelson, L. K. (2012). The First Slow Step: Differential Effects of Object and Word-Form Familiarization on Retention of Fast-Mapped Words. *Infancy : The Official Journal of the International Society on Infant Studies*, *17*(3), 295. https://doi.org/10.1111/J.1532-7078.2011.00081.X

LaTourrette, A., Waxman, S., Wakschlag, L. S., Norton, E. S., & Weisleder, A. (2023). From recognizing known words to learning new ones: Comparing online speech processing in typically developing and late-talking 2-year-olds. *Journal of Speech, Language, and Hearing Research*, *66*(5), 1658–1677. https://doi.org/10.1044/2023_JSLHR-22-00580;WGROUP:STRING:ASHA

Leung, A., Tunkel, A., & Yurovsky, D. (2021). Parents Fine-Tune Their Speech to Children's Vocabulary Knowledge. *Psychological Science*, *32*(7), 975–984. https://doi.org/10.1177/0956797621993104/ASSET/IMAGES/LARGE/10.1177_0956797621993104-FIG4.JPEG

Lorang, E., Hanania, A., & Venker, C. E. (2023). Parent Certainty and Consistency When Completing Vocabulary Checklists in Young Autistic Children. *Journal of Speech, Language, and Hearing Research*, 1–16. https://doi.org/10.1044/2023_JSLHR-22-00623

MacWhinney, B. (2000). The Childes Project : Tools for Analyzing Talk, Volume II: the Database. *The Childes Project*. https://doi.org/10.4324/9781315805641

Maital, S. L., Dromi, E., Sagi, A., & Bornstein, M. H. (2000). The Hebrew Communicative Development Inventory: language specific properties and cross-linguistic generalizations. *Journal of Child Language*, *27*(1), 43–67. https://doi.org/10.1017/S0305000999004006

Mani, N., & Plunkett, K. (2011). Phonological priming and cohort effects in toddlers. *Cognition*, *121*(2), 196–206. https://doi.org/10.1016/J.COGNITION.2011.06.013

Marchman, V. A., Dale, P. S., & Larry, F. (2023). *MacArthur-Bates Communicative Development Inventories User's Guide and Technical Manual* (Third). Brookes Publishing Co. https://products.brookespublishing.com/MacArthur-Bates-Communicative-Development-Inventories-Users-Guide-and-Technical-Manual-Third-Edition-P1481.aspx

Marchman, V. A., & Fernald, A. (2008a). Speed of word recognition and vocabulary knowledge in infancy predict cognitive and language outcomes in later childhood. *Developmental Science*, *11*(3), F9. https://doi.org/10.1111/J.1467-7687.2008.00671.X

Marchman, V. A., & Fernald, A. (2008b). Speed of word recognition and vocabulary knowledge in infancy predict cognitive and language outcomes in later childhood. *Developmental Science*, *11*(3), F9–F16. https://doi.org/10.1111/J.1467-7687.2008.00671.X

Markman, E. M. (1990). Constraints Children Place on Word Meanings. *Cognitive Science*, *14*(1), 57–77. https://doi.org/10.1207/S15516709COG1401_4

Markman, E. M., & Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology*, *20*(2), 121–157. https://doi.org/10.1016/0010-0285(88)90017-5

Masterson, J., Druks, J., & Gallienne, D. (2008). Object and action picture naming in three- and five-year-old children*. *Journal of Child Language*, *35*(2), 373–402. https://doi.org/10.1017/S0305000907008549

Mayor, J., & Plunkett, K. (2011). A statistical estimate of infant and toddler vocabulary size from CDI analysis. *Developmental Science*, *14*(4), 769–785. https://doi.org/10.1111/J.1467-7687.2010.01024.X

Mayor, J., & Plunkett, K. (2014). Shared understanding and idiosyncratic expression in early vocabularies. *Developmental Science*, *17*(3), 412–423. https://doi.org/10.1111/DESC.12130

McNamara, T. P., & Sternberg, R. J. (1983). Mental models of word meaning. *Journal of Verbal Learning and Verbal Behavior*, *22*(4), 449–474. https://doi.org/10.1016/S0022-5371(83)90291-8

McRae, K., De Sa, V. R., & Seidenberg, M. S. (1997). On the Nature and Scope of Featural Representations of Word Meaning. *Journal of Experimental Psychology: General*, *126*(2), 99–130. https://doi.org/10.1037/0096-3445.126.2.99

Meints, K., Plunkett, K., & Harris, P. L. (1999). When does an ostrich become a bird? The role of typicality in early word comprehension. *Developmental Psychology*, *35*(4), 1072–1078. https://doi.org/10.1037/0012-1649.35.4.1072

Mervis, C. B., & Rosch, E. (1981). Categorization of Natural Objects. *Annual Review Psychology*, *32*, 89–115.

Migo, E. M., Montaldi, D., & Mayes, A. R. (2013). A visual object stimulus database with standardized similarity information. *Behavior Research Methods*, *45*(2), 344–354. https://doi.org/10.3758/S13428-012-0255-4/FIGURES/6

Miller, L. E., Perkins, K. A., Dai, Y. G., & Fein, D. A. (2017). Comparison of parent report and direct assessment of child skills in toddlers. *Research in Autism Spectrum Disorders*, *41–42*, 57–65. https://doi.org/10.1016/J.RASD.2017.08.002

Moore, C., & Bergelson, E. (2024). Wordform variability in infants' language environment and its effects on early word learning. *Cognition*, *245*, 105694.

Muradoglu, M., Cimpian, J. R., & Cimpian, A. (2023). Mixed-Effects Models for Cognitive Development Researchers. *Journal of Cognition and Development*, *24*(3), 307–340. https://doi.org/10.1080/15248372.2023.2176856

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. https://doi.org/10.1016/J.JECP.2017.04.017

Nikolaus, M., & Fourtassi, A. (2021). Evaluating the acquisition of semantic knowledge from cross-situational learning in artificial neural networks. In *Proceedings of the workshop on cognitive modeling and computational linguistics* (pp. 200-210). https://doi.org/10.18653/v1/2021.cmcl-1.24ï

Nomikou, I., Rohlfing, K. J., Cimiano, P., & Mandler, J. M. (2019). Evidence for Early Comprehension of Action Verbs. *Language Learning and Development*, *15*(1), 64–74. https://doi.org/10.1080/15475441.2018.1520639;SUBPAGE:STRING:ABSTRACT;WGROUP:STRING:PUBLICATION

Olson, R. H., Pomper, R., Potter, C. E., Hay, J. F., Saffran, J. R., Ellis Weismer, S., & Lew-Williams, C. (2020). *peyeCoder: An open-source program for coding eye movements*.

Peer, E., Rothschild, D., Gordon, A., Evernden, Z., & Damer, E. (2022). Data quality of platforms and panels for online behavioral research. *Behavior Research Methods*, *54*(4), 1643–1662. https://doi.org/10.3758/S13428-021-01694-3/TABLES/13

Perry, L. K., Kucker, S. C., Horst, J. S., & Samuelson, L. K. (2022). Late bloomer or language disorder? Differences in toddler vocabulary composition associated with long-term language outcomes. *Developmental Science*, *26*(4), e13342. https://doi.org/10.1111/DESC.13342;PAGE:STRING:ARTICLE/CHAPTER

Perry, L. K., & Saffran, J. R. (2017). Is a Pink Cow Still a Cow? Individual Differences in Toddlers' Vocabulary Knowledge and Lexical Representations. *Cognitive Science*, *41*(4), 1090–1105. https://doi.org/10.1111/COGS.12370

Perry, L. K., & Samuelson, L. K. (2011). The shape of the vocabulary predicts the shape of the bias. *Frontiers in Psychology*, *2*(NOV), 16408. https://doi.org/10.3389/FPSYG.2011.00345/BIBTEX

Perry, L. K., Samuelson, L. K., & Burdinie, J. B. (2014). Highchair philosophers: the impact of seating context-dependent exploration on children's naming biases. *Developmental Science*, *17*(5), 757–765. https://doi.org/10.1111/DESC.12147

Perry, L. K., Samuelson, L. K., Malloy, L. M., & Schiffer, R. N. (2010). Learn locally, think globally: Exemplar variability supports higher-order generalization and word learning. *Psychological Science*, *21*(12), 1894–1902.

https://doi.org/10.1177/0956797610389189/ASSET/IMAGES/LARGE/10.1177_0956797610389189-FIG4.JPEG

Peter, M. S., Durrant, S., Jessop, A., Bidgood, A., Pine, J. M., & Rowland, C. F. (2019). Does speed of processing or vocabulary size predict later language growth in toddlers? *Cognitive Psychology*, *115*, 101238. https://doi.org/10.1016/J.COGPSYCH.2019.101238

Peters, R., & Borovsky, A. (2019). Modeling early lexico-semantic network development: Perceptual features matter most. *Journal of Experimental Psychology: General*, *148*(4), 763–782. https://doi.org/10.1037/xge0000596

Pomiechowska, B., & Gliga, T. (2019). Lexical Acquisition Through Category Matching: 12-Month-Old Infants Associate Words to Visual Categories. *Psychological Science*, *30*(2), 288–299. https://doi.org/10.1177/0956797618817506/ASSET/IMAGES/LARGE/10.1177_0956797618817506-FIG4.JPEG

Pompéia, S., Miranda, M. C., & Bueno, O. F. A. (2001). A set of 400 pictures standardised for Portuguese: norms for name agreement, familiarity and visual complexity for children and adults. *Arquivos de Neuro-Psiquiatria*, *59*(2 B), 330–337. https://doi.org/10.1590/S0004-282X2001000300004

Pomper, R., & Saffran, J. R. (2019). Familiar Object Salience Affects Novel Word Learning. *Child Development*, *90*(2), e246–e262. https://doi.org/10.1111/CDEV.13053

Potter, C. E., & Lew-Williams, C. (2023). Frequent vs. infrequent words shape toddlers' real-time sentence comprehension. *Journal of Child Language*, 1–11. https://doi.org/10.1017/S0305000923000387

Quine, W. V. O. (1960). *Word and Object*. MIT Press.

R Core Team. (2024). *R: A language and environment for statistical computing (Version 4.5.0)*. R Foundation for Statistical Computing.

Ramsey, R. E. (2022). Individual differences in word senses. *Cognitive Linguistics*, *33*(1), 65–93. https://doi.org/10.1515/COG-2021-0020/DOWNLOADASSET/SUPPL/J_COG-2021-0020_SUPPL.ZIP

Raudenbush, S. W. (2009). Adaptive Centering with Random Effects: An Alternative to the Fixed Effects Model for Studying Time-Varying Treatments in School Settings. *Education Finance and Policy*, *4*(4), 468–491. https://doi.org/10.1162/EDFP.2009.4.4.468

Rescorla, L. (1989). The Language Development Survey. *Journal of Speech and Hearing Disorders*, *54*(4), 587–599. https://doi.org/10.1044/JSHD.5404.587

Rescorla, L. (2002). Language and reading outcomes to age 9 in late-talking toddlers. *Journal of Speech, Language, and Hearing Research*, *45*(2), 360–371. https://doi.org/10.1044/1092-4388(2002/028)

Rescorla, L. (2011). Late talkers: Do good predictors of outcome exist? *Developmental Disabilities Research Reviews*, *17*(2), 141–150. https://doi.org/10.1002/DDRR.1108,

Ring, E. D., & Fenson, L. (2000). The correspondence between parent report and child performance for receptive and expressive vocabulary beyond infancy. *First Language*, *20*(20), 141–159. https://doi.org/10.1177/014272370002005902

Roberts, S. O., Bareket-Shavit, C., Dollins, F. A., Goldie, P. D., & Mortenson, E. (2020). Racial Inequality in Psychological Research: Trends of the Past and Recommendations for the Future. *Perspectives on Psychological Science*, *15*(6), 1295–1309. https://doi.org/10.1177/1745691620927709/ASSET/163B879A-9677-4179-90B5-0ED21F22739B/ASSETS/IMAGES/LARGE/10.1177_1745691620927709-FIG4.JPG

Rodd, J. M. (2020). Settling Into Semantic Space: An Ambiguity-Focused Account of Word-Meaning Access. *Perspectives on Psychological Science*, *15*(2), 411–427. https://doi.org/10.1177/1745691619885860/ASSET/IMAGES/LARGE/10.1177_1745691619885860-FIG1.JPEG

Rodd, J. M., Cai, Z. G., Betts, H. N., Hanby, B., Hutchinson, C., & Adler, A. (2016). The impact of recent and long-term experience on access to word meanings: Evidence from large-scale internet-based experiments. *Journal of Memory and Language*, *87*, 16–37. https://doi.org/10.1016/J.JML.2015.10.006

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, *7*(4), 573–605. https://doi.org/10.1016/0010-0285(75)90024-9

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*(3), 382–439. https://doi.org/10.1016/0010-0285(76)90013-X

Roy, B. C., Frank, M. C., DeCamp, P., Miller, M., & Roy, D. (2015). Predicting the birth of a spoken word. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(41), 12663–12668. https://doi.org/10.1073/pnas.1419773112

Royston, P., Altman, D. G., & Sauerbrei, W. (2006). Dichotomizing continuous predictors in multiple regression: a bad idea. *Statistics in Medicine*, *25*(1), 127–141. https://doi.org/10.1002/SIM.2331

Saffran, J. R., & Kirkham, N. Z. (2018). Infant Statistical Learning. *Annual Review of Psychology*, *69*(1), 181–203. https://doi.org/10.1146/ANNUREV-PSYCH-122216-011805/CITE/REFWORKS

Samuelson, L. K., & McMurray, B. (2017). What does it take to learn a word? *Wiley Interdisciplinary Reviews: Cognitive Science*, *8*(1–2), e1421. https://doi.org/10.1002/WCS.1421

Samuelson, L. K., & Smith, L. B. (1999). Early noun vocabularies: do ontology, category structure and syntax correspond? *Cognition*, *73*(1), 1–33. https://doi.org/10.1016/S0010-0277(99)00034-7

Sanchez, A., Meylan, S. C., Braginsky, M., MacDonald, K. E., Yurovsky, D., & Frank, M. C. (2019). childes-db: A flexible and reproducible interface to the child language data exchange system. *Behavior Research Methods*, *51*(4), 1928–1941. https://doi.org/10.3758/S13428-018-1176-7/FIGURES/8

Slone, L. K., Abney, D. H., Smith, L. B., & Yu, C. (2023). The temporal structure of parent talk to toddlers about objects. *Cognition*, *230*, 105266. https://doi.org/10.1016/J.COGNITION.2022.105266

Slone, L. K., Smith, L. B., & Yu, C. (2019). Self-generated variability in object images predicts vocabulary growth. *Developmental Science*, *22*(6). https://doi.org/10.1111/desc.12816

Smith, L., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, *106*(3), 1558. https://doi.org/10.1016/J.COGNITION.2007.06.010

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(2), 174–215. https://doi.org/10.1037/0278-7393.6.2.174

Sommerfeld, L., Staudte, M., & Kray, J. (2022). Ratings of name agreement and semantic categorization of 247 colored clipart pictures by young German children. *Acta Psychologica*, *226*, 103558. https://doi.org/10.1016/J.ACTPSY.2022.103558

Southgate, V., & Meints, K. (2000). Typicality, Naming, and Category Membership in Young Children. *Cognitive Linguistics*, *11*(1–2), 5–16. https://doi.org/10.1515/COGL.2001.011

Srinivasan, M., & Rabagliati, H. (2021). The Implications of Polysemy for Theories of Word Learning. *Child Development Perspectives*, *15*(3), 148–153. https://doi.org/10.1111/CDEP.12411;CTYPE:STRING:JOURNAL

Steil, J. N., Friedrich, C. K., & Schild, U. (2021). No Evidence of Robust Noun-Referent Associations in German-Learning 6- to 14-Month-Olds. *Frontiers in Psychology*, *12*, 718742. https://doi.org/10.3389/FPSYG.2021.718742/BIBTEX

Styles, S., & Plunkett, K. (2009). What is ’word understanding ’ for the parent of a one-year-old ? Matching the difficulty of a lexical comprehension task to parental CDI report. *Journal of Child Language*, *36*(4), 895–908. https://doi.org/10.1017/S0305000908009264

Suanda, S. H., Barnhart, M., Smith, L. B., & Yu, C. (2019). The Signal in the Noise: The Visual Ecology of Parents’ Object Naming. *Infancy*, *24*(3), 455–476. https://doi.org/10.1111/INFA.12278

Suanda, S. H., Smith, L. B., & Yu, C. (2017). The Multisensory Nature of Verbal Discourse in Parent–Toddler Interactions. In *Multisensory Perception and Communication* (pp. 56-73). Routledge. https://doi.org/10.1080/87565641.2016.1256403

Suarez-Rivera, C., Linn, E., & Tamis-LeMonda, C. S. (2022). From Play to Language: Infants' Actions on Objects Cascade to Word Learning. *Language Learning*, *72*(4), 1092–1127. https://doi.org/10.1111/LANG.12512

Swingley, D., & Aslin, R. N. (2000). Spoken word recognition and lexical representation in very young children. *Cognition*, *76*(2), 147–166. https://doi.org/10.1016/S0010-0277(00)00081-0

Swingley, D., Pinto, J. P., & Fernald, A. (1999). Continuous processing in word recognition at 24 months. *Cognition*, *71*(2), 73–108. https://doi.org/10.1016/S0010-0277(99)00021-9

Syrnyk, C., & Meints, K. (2017). Bye-bye mummy – Word comprehension in 9-month-old infants. *British Journal of Developmental Psychology*, *35*(2), 202–217. https://doi.org/10.1111/BJDP.12157

Tamis-Lemonda, C. S., Custode, S., Kuchirko, Y., Escobar, K., & Lo, T. (2018). Routine language: Speech directed to infants during home activities. *Child development, 90*(6), 2135-2152. https://doi.org/10.1111/cdev.13089

Tamis-LeMonda, C. S., Kuchirko, Y., & Song, L. (2014). Why Is Infant Language Learning Facilitated by Parental Responsiveness? *Current Directions in Psychological Science*, *23*(2), 121-126. https://doi.org/10.1177/0963721414522813

Tardif, T., Shatz, M., & Naigles, L. (1997). Caregiver speech and children's use of nouns versus verbs: A comparison of English, Italian, and Mandarin. *Journal of Child Language*, *24*(3), 535–565. https://doi.org/10.1017/S030500099700319X

Thompson, B., Roberts, S. G., & Lupyan, G. (2020). Cultural influences on word meanings revealed through large-scale semantic alignment. *Nature Human Behaviour*, *4*(10), 1029–1038. https://doi.org/10.1038/s41562-020-0924-8

Tincoff, R., & Jusczyk, P. W. (1999). Some beginnings of word comprehension in 6-month-olds. *Psychological Science*, *10*(2), 172–175.

U.S. Census Bureau. (2020). Hispanic or Latino, and Not Hispanic or Latino by Race. In *Decennial Census*.

U.S. Census Bureau. (2023a). Educational Attainment. In *American Community Survey*.

U.S. Census Bureau. (2023b). Income in the Past 12 Months (in 2023 Inflation-Adjusted Dollars). In *American Community Survey*.

Valleau, M. J., Konishi, H., Golinkoff, R. M., Hirsh-Pasek, K., & Arunachalam, S. (2018). An Eye-Tracking Study of Receptive Verb Knowledge in Toddlers. *Journal of Speech,* *Language, and Hearing Research*, *61*(12), 2917–2933. https://doi.org/10.1044/2018_JSLHR-L-17-0363

Venker, C. E., Haebig, E., Edwards, J., Saffran, J. R., & Ellis Weismer, S. (2016). Brief Report: Early Lexical Comprehension in Young Children with ASD: Comparing Eye-Gaze Methodology and Parent Report. *Journal of Autism and Developmental Disorders*, *46*(6), 2260–2266. https://doi.org/10.1007/S10803-016-2747-Z/FIGURES/2

Vlach, H. A., & DeBrock, C. A. (2019). Statistics learned are statistics forgotten: Children's retention and retrieval of cross-situational word learning. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *45*(4), 700. https://doi.org/10.1037/XLM0000611

Wang, X., & Bi, Y. (2021). Idiosyncratic Tower of Babel: Individual Differences in Word-Meaning Representation Increase as Word Abstractness Increases. *Psychological Science*, *32*(10), 1617–1635. https://doi.org/10.1177/09567976211003877/ASSET/IMAGES/LARGE/10.1177_09567976211003877-FIG4.JPEG

Weaver, H., & Saffran, J. (under review). Interrogating early word knowledge: Factors that influence the alignment between caregiver report and experimental measures.

Weaver, H., Zettersten, M., & Saffran, J. R. (2024). Becoming word meaning experts: Infants' processing of familiar words in the context of typical and atypical exemplars. *Child Development*, *95*(5), e352–e372. https://doi.org/10.1111/CDEV.14120

Wiley, J., George, T., & Rayner, K. (2018). Baseball fans don't like lumpy batters: Influence of domain knowledge on the access of subordinate meanings. *Quarterly Journal of Experimental Psychology*, *71*(1), 93–102. https://doi.org/10.1080/17470218.2016.1251470

Wojcik, E. H., Pierce, M. C., Stevens, G., & Goulding, S. J. (2024). Referent-oriented interactions in infancy: A naturalistic, longitudinal case study from an English-speaking household. *Infant Behavior and Development*, *74*, 101911. https://doi.org/10.1016/J.INFBEH.2023.101911

Wojcik, E. H., Zettersten, M., Viridiana, |, & Benitez, L. (2022). The map trap: Why and how word learning research should move beyond mapping. *Wiley Interdisciplinary Reviews: Cognitive Science, 13*(4), e1596. https://doi.org/10.1002/wcs.1596

Wulff, D. U., Hills, T. T., & Mata, R. (2022). Structural differences in the semantic networks of younger and older adults. *Scientific Reports, 12(1*), 1–13. https://doi.org/10.1038/s41598-022-11698-4

Yu, C., Smith, L. B., & Pereira, A. F. (2008). Grounding Word Learning in Multimodal Sensorimotor Interaction. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *30*, 30.

Yu, C., Suanda, S. H., & Smith, L. B. (2019). Infant sustained attention but not joint attention to objects at 9 months predicts vocabulary at 12 and 15 months. *Developmental Science*, *22*(1). https://doi.org/10.1111/desc.12735

Zarcone, A., & Demberg, V. (2021). Interaction of Script Knowledge and Temporal Discourse Cues in a Visual World Study. *Discourse Processes*, *58*(9), 804–819. https://doi.org/10.1080/0163853X.2021.1930807;WGROUP:STRING:PUBLICATION

Zuniga-Montanez, C., & Krott, A. (2025). Late Talkers can generalise trained labels by object shape similarities, but not unfamiliar labels. *Journal of Child Language*, *52*(4), 815–838. https://doi.org/10.1017/S0305000924000163