COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 49 (2025) e70104

ISSN: 1551-6709 online
DOI: 10.1111/cogs.70104

Learning Partial Word Meanings From Referentially Ambiguous Naming Events

Nina Schoener,[a] Sara C. Johnson,[b] Sumarga H. Suanda[b,c]

[a]*Department of Psychology, University of California, Berkeley*
[b]*Department of Psychological Sciences, University of Connecticut*
[c]*Institute for Brain and Cognitive Sciences, University of Connecticut*

Received 19 October 2024; received in revised form 7 July 2025; accepted 22 July 2025

Abstract

Both classic thought experiments and recent empirical evidence suggest that children frequently encounter new words whose meanings are underdetermined by the extralinguistic contexts in which they occur. The role that these referentially ambiguous events play in children's word learning is central to ongoing debates in the field. Do children learn words from referentially ambiguous events via an incremental learning process? Or, do children learn words primarily from the rare referentially transparent events they experience? Across two experiments with adults as model word learners, the current work asks whether the answer to these questions depends in part on how word learning is assessed. Participants were asked to learn the meanings of novel words solely from their referentially ambiguous contexts. When learning was assessed by asking participants to identify the exact meanings of those novel words, participants struggled mightily. However, when learning was assessed by asking the same participants to identify which of two new contexts the novel word most likely occurred in, even those who failed the exact meaning assessment succeeded. These data suggest that although referentially ambiguous events may fall short in allowing learners to identify a word's exact meaning, they nevertheless lead learners into the right regions of semantic space. These findings are a reminder of the pervasiveness of partial word learning effects in vocabulary acquisition and highlight that the resolution to the debate over the role of referentially ambiguous events in learning may depend on how learning is defined.

Keywords: Cross-situational word learning; Referential ambiguity; Partial knowledge

Correspondence should be sent to Sumarga H. Suanda, Department of Psychological Sciences, University of Connecticut, 406 Babbidge Road, Unit 1020, Storrs, CT 06269-1020. E-mail: s.suanda@uconn.edu

1. Introduction

Young children encounter new words in a wide variety of situations. Sometimes, these situations render the words' meanings transparent by virtue of the extralinguistic context. For example, if a parent uttered to their toddler, "Would you like this *apple*" as they handed them an apple, there is relatively little ambiguity as to what the word "apple" refers to. Often, however, this is not the case. For example, imagine a parent uttering to their toddler, "Let's not forget to buy *apples*" as they entered the produce section of a grocery store. In this scenario, there is much more ambiguity as to what the word "apple" refers to: It could refer to any one of the many fruits or vegetables on display, it could refer to a superordinate category term (e.g., "fruit"), or it could even refer to something that is not on display. Most word learning scholars agree that highly referentially transparent events, like the one depicted in the first scenario, can do a lot to boost learning (Cartmill et al., 2013; Horst & Samuelson, 2008; Tamis-LeMonda, Kuchirko, & Song, 2014; Yu & Smith, 2012b). There is much less agreement, however, on the part that referentially ambiguous events, like the one depicted in the second example, play in the learning process. Whereas some have argued that referentially ambiguous events do contribute to word learning (Yu & Smith, 2007; Yurovsky, Smith, & Yu, 2013; Zhang, Yurovsky, & Yu, 2021), others have questioned whether they move the learning needle at all (Gleitman & Trueswell, 2020; Medina, Snedeker, Trueswell, & Gleitman, 2011; Trueswell et al., 2016).

The current study asks whether the contributions of referentially ambiguous events to word learning depend in part on how word learning is defined. That is, if the definition of learning is constrained to mapping words onto their precise meanings, then perhaps referentially ambiguous events, as some have shown (Medina et al., 2011; LaTourrette et al., 2022), may indeed fall short in contributing to this process. Importantly however, many have argued, and demonstrated empirically, that learning the meaning of a word is often not an all-or-none process (Carey, 2010; Clark, 2023; Durso & Shore, 1991; Yurovsky, Fricker, Yu, & Smith, 2014) and that it is not uncommon for word learners of all ages to know some aspects of a word's meaning without knowing its exact meaning (Ameel, Malt, & Storms, 2008; Hendrickson, Mitsven, Poulin-Dubois, Zesiger, & Friend, 2015; Kominsky & Keil, 2014; Lo, Rosslund, Chai, Mayor, & Kartushina, 2021). Motivated by such demonstrations of partial word knowledge during vocabulary acquisition, the current study revisits the role that referentially ambiguous naming events play in word learning by probing the nature of learners' knowledge when they fail to acquire a word's precise meaning from those events. Of particular interest is whether referentially ambiguous events equip learners with partial information that lands them in the correct regions of semantic space even when it does not get them to the correct coordinates within that space. Returning to the referentially ambiguous event illustrated in the previous paragraph, although that hypothetical toddler might struggle to figure out exactly to what the word "apple" refers, they might nevertheless gain valuable knowledge like that apples are likely a type of food, perhaps even a type of produce, and likely something one can find in a grocery store. This is the type of partial word learning probed in the current study.

1.1. Word learning from referential ambiguity naming events

Referentially ambiguous naming events and their place in early word learning have been of central interest to language development researchers for over 50 years (see MacNamara, 1972). Early research on this topic focused largely on the fact that all naming events are ambiguous to some extent (for discussion, see Swingley, 2010). That is, as many word learning experimentalists have pointed out over the years, even in the seemingly referentially transparent event mentioned above, the child must still figure out that the word "apple" goes with the meaning *apple* and not with the meanings *fruit*, *red delicious*, *stem*, *apple slice*, and the myriad other candidate meanings. Over the years, countless experimental studies were devoted to understanding how there could be rampant ambiguity in the input on the one hand and rapid vocabulary growth in children on the other. What this work revealed is that children possess an arsenal of conceptual, linguistic, and socio-cognitive tools that allow them to interpret naming events fairly accurately at the moment those events are encountered (for reviews, see Bloom, 2000; Golinkoff & Hirsh-Pasek, 2000; Woodward & Markman, 1998).

In more recent years, researchers have expanded their focus to not only how children overcome the ambiguity present in fairly simple naming events (i.e., the parent saying "apple" while handing the toddler an apple) but also how children overcome the ambiguity in more complex events (i.e., the parent saying "apple" in the produce section with many candidate referents). One proposal is that children could overcome that ambiguity across multiple encounters with that word. That is, as long as words co-occur more frequently with their referents than with the referents of other words, then a learner capable of tracking word-referent co-occurrence statistics could learn from individually ambiguous naming events. In a pair of foundational experiments, Yu and Smith asked whether adult and infant learners are capable of such cross-situational statistical word learning (Smith & Yu, 2008; Yu & Smith, 2007). In these studies, participants were presented with naming events where they heard multiple words and saw multiple objects. Importantly, each individual event was ambiguous in that there was no information that was revealing as to which words went with which objects. Across events, however, word–object pairs reappeared in different configurations (e.g., word–object pair "*a*-A" might co-occur with word–object pair "*b*-B" on Trial 1 but with word–object pair "*c*-C" on Trial 5). As a result, although there was referential ambiguity *within* events, there was referential coherence *across* events. What Yu and Smith found, as well as many others since, is that a wide range of learners, including infants, young children, and adults, can learn words reliably from this type of ambiguous input (for reviews of this literature, see Roembke, Simonetti, Koch, & Philipp, 2023; Smith, Suanda, & Yu, 2014; Zhang, Chen, & Yu, 2019).

1.2. The challenge of real-world referential ambiguity

Importantly, much of the work on word learning from referentially ambiguous events comes from highly simplified laboratory experiments. As a result, the relevance of these findings to everyday learning experiences and to real-world levels of ambiguity has been called into question (Gleitman & Trueswell, 2020; Medina et al., 2011). In one set of studies employing a methodology known as the *Human Simulation Paradigm*, Medina et al. (2011)

asked whether even mature adult learners were capable of identifying word meanings from everyday referentially ambiguous naming events. In their study, Medina and colleagues captured audio-visual recordings of caregiver–child interactions and then created short video vignettes around the moments in which caregivers uttered target words of interest (e.g., "ball"). These vignettes were edited to be muted except for a beep that was inserted at the precise moment the target word was uttered. The researchers then presented these vignettes to naïve adult observers, asking them to guess the identity of the target word. One of the striking findings from this study was the degree to which even adults, with their mature cognitive and socio-communicative understanding, struggled to identify the correct word. That is, even for common nouns, only 17% of observers correctly identified the target word (see also Trueswell et al., 2016). Moreover, in a critical follow-up study with a cross-situational design, Medina and colleagues found that providing participants with a set of multiple referentially ambiguous vignettes belonging to the same target word did not improve their word identification performance. Word identification was only improved when a referentially *transparent* vignette was included in the set. Together, Medina and colleagues' findings (see also LaTourrette et al., 2022) suggest a very different perspective on the role referentially ambiguous events play in word learning. Namely, that their contribution to learning is minimal and that much of word learning takes place not from aggregating across referentially ambiguous events but rather from rapidly learning via the few referentially transparent naming events that children do encounter (see Gleitman & Trueswell, 2020).

1.3. The debate over learning from referentially ambiguous events

Medina and colleagues' study by no means settled the debate over whether referentially ambiguous events contribute to learning. In one notable follow-up, Yurovsky and colleagues asked whether the perspective from which referentially ambiguous events are viewed affects our estimates of their contributions to learning. In their study, audio-visual recordings of naming events were captured from both a third-person perspective, as had been done by Medina and colleagues, and a first-person perspective, via toddler-worn head-mounted cameras (see Smith, Jayaraman, Clerkin, & Yu, 2018; Smith, Yu, & Pereira, 2011; Yoshida & Smith, 2008). In a study modeled after Medina and colleagues' cross-situational learning study, Yurovsky and colleagues asked adult learners to identify the target referent shared by a series of referentially ambiguous naming event vignettes. Critically, participants were presented either with referentially ambiguous third-person perspective vignettes or referentially ambiguous first-person perspective vignettes. What Yurovsky and colleagues found was that whereas learning from third-person perspective vignettes did not improve cross-situationally, learning from first-person perspective vignettes did (Yurovsky et al., 2013). These results suggest the possibility that not all forms of referential ambiguity are created equal and that ambiguous naming events when viewed from the first-person learner's perspective can contribute to successful word learning (see also Zhang et al., 2021).

1.4. The motivation behind the current study

Like Yurovsky and colleagues' work, the current study seeks to deepen our understanding of the role that referentially ambiguous naming events play in word learning. In contrast, however, to Yurovsky and colleagues' focus on the *input* side of the learning equation (i.e., the visual perspective from which referentially ambiguous events are viewed), the current work focuses on the *output* side of the equation or on how learning is assessed and defined. That is, in Medina and colleagues' Human Simulation Paradigm study that demonstrated minimal learning via referentially ambiguous events, the key dependent measure of learning was the extent to which learners successfully guessed the exact identity of the word (see also Cartmill et al., 2013; Gillette, Gleitman, Gleitman, & Lederer, 1999; Snedeker & Gleitman, 2004). For example, had the novel word "vash" been paired with a series of referentially ambiguous vignettes in which the caregivers had uttered the word "apple," then participants were only considered to have learned the word "vash" if they guessed that this word meant the word "apple"; had participants guessed semantically related words (e.g., "pear," "fruit," "tree," "food," etc.), they would not have been credited with learning. There is certainly validity to employing this stringent metric of learning because, after all, the word "apple" does not mean *pear*, *fruit*, *tree*, or other *apple*-adjacent meanings. However, this stringent metric of learning does not take into account two important details about everyday word learning and how it is assessed. First, infants, toddlers, children, and even adults, often go through extended stages of partial or fragmented word knowledge in which they know some aspects of a word's meaning while falling short of knowing the word's "full" meaning (Ameel et al., 2008; Hendrickson et al., 2015; Kominsky & Keil, 2014; Lo et al., 2021; O'Rear, McNeil, & Kirkland, 2020; Tillman & Barner, 2015; Wagner, Dobkins, & Barner, 2013). For example, Ameel et al. (2008) have shown that the meanings of even some "early-learned" common nouns like "bottle" do not converge onto adult-like meanings until early adolescence. Thus, one possibility is that referentially ambiguous events contribute to word learning by shaping partial and imperfect word knowledge without being the key ingredient that propels learners to full word knowledge. Second, and importantly, if one considers the common methods by which clinicians and researchers credit children for knowing a word (e.g., parent report checklists, preferential looking tasks, word prompt picture selection tasks), these methods have a much less stringent threshold for success than Medina and colleagues' exact meaning threshold. Indeed, as Bloom once noted, the more common perspective in the field is that children have learned a word "when they have a *rough idea* [emphasis added] as to what the word refers to" (Bloom, 2001, p. 1129). Thus, to the extent that partial word knowledge, at least in many contexts, is what is meant by "word learning," then if referentially ambiguous events lead to such knowledge, it should follow that they be taken seriously as valuable contributors to word learning.

1.5. Current study overview

The two experiments below test the idea that referentially ambiguous naming events are sufficient to yield valuable partial word knowledge. In both experiments, adult word learners were asked to identify the meanings of novel words (e.g., "blicket") from a series of referentially ambiguous naming events involving those words. Their learning of these novel words was assessed in two ways. First, learning was assessed via a word identity test that tapped into whether learners successfully identified the *exact* meaning of those novel words (each novel word's meaning corresponded to the meaning of an English word like "apple"). Second,

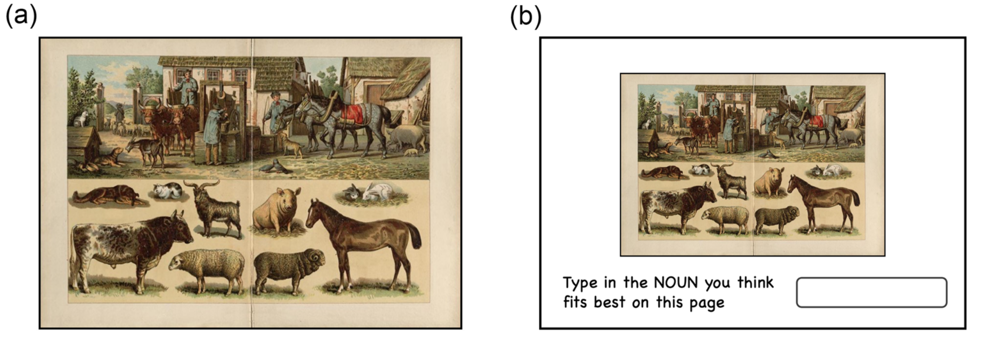

Fig. 1. Sample referentially ambiguous vignette.
Note. This figure depicts a sample referentially ambiguous event for the target noun "dog" (a) along with how that vignette appeared in the norming study (b). Due to copyright issues, we are unable to depict the actual stimuli used in this study. This scene, which is representative of the ambiguity level of the vignettes used in this study, comes from Frederick Warne's "*The Children's Object Book*," which is in the public domain (see Frederick Warne, n.d.).

learning was also assessed via a two-alternative forced choice (2AFC) test in which learners simply had to indicate which of two previously unseen naming events more likely involved the novel word. This dual assessment allowed us to test the idea that although a word's referentially ambiguous naming events may be insufficiently informative to lead learners to that word's exact meaning, these events may nevertheless be sufficient to lead learners to the right semantic ballpark. That is, although referentially ambiguous scenes of the word "apple" may not saliently depict the meaning *apple*, they may nevertheless depict a set of contexts (e.g., stores, kitchens, orchards), objects (e.g., foods, bowls, trees), or activities (e.g., meals, picnics, shopping) that help sketch the broader semantic space in which *apple* resides. If this idea is correct, then even participants who fail in the word identity assessment that requires exact knowledge of word meaning might succeed at the 2AFC assessment that requires only partial knowledge of word meaning. This is the key hypothesis tested in these two experiments, which differ only in the level of referential ambiguity in the naming events that learners faced.

2. Experiment 1

Experiment 1 served as an initial test of whether referentially ambiguous naming events could support the learning of partial word knowledge. In this experiment, adult learners saw a series of referentially ambiguous vignettes belonging to an English word (e.g., "apple"). The vignettes were pages, edited to be textless, of children's picture books in which the English word had originally appeared in the page's text (see Fig. 1a). Importantly, each vignette had been determined via prior norming to be referentially ambiguous. It is worth briefly noting here that the use of picture book vignettes is a departure from previous research that has utilized video-based vignettes (see Medina et al., 2011; Yurovsky et al., 2013; Zhang et al., 2021). Although we discuss the implications of this methodological choice in more detail in the *General Discussion*, it is important to keep in mind that picture books, and the words within them, do represent a type of real-world language input for many children (Horst & Houston-Price, 2015; Montag, Jones, & Smith, 2015; Nation, Dawson, & Hsiao, 2022). The referentially ambiguous vignettes were paired with novel words (e.g., "blicket") and participants' task was to learn the meanings of those words.

In a series of test trials, participants were first asked to guess the English word that corresponded to that novel word. Participants were then asked to choose which of two previously unseen vignettes (one of which had also contained the novel word's English translation in its original text) was more likely to contain the novel word. For example, had the novel word "blicket" corresponded to the English word "apple," participants were asked to select which of two scenes most likely contained the novel word "blicket": a new scene that had contained the word "apple" in its original text or a new scene that had not contained "apple" in its original text. We hypothesized that like Medina et al. (2011) participants, adult learners in this study would struggle in the exact meaning word identity assessment. Of particular interest, however, is whether this exact meaning assessment masks partial word knowledge that learners can successfully extract from referentially ambiguous events. Thus, the primary analyses of interest in this study are whether participants who failed in the word identity assessment nevertheless succeeded in the alternative forced choice assessment.

2.1. Methods

2.1.1. Participants

Using G*Power 3 (Faul, Erdfelder, Lang, & Buchner, 2007), an a priori power analysis was conducted to determine the required sample size to detect a near-moderate effect size for a one-tailed single-samples *t*-test (our primary analyses of interest entail comparisons to a chance baseline) at an alpha-level of .01 and with a power of 80%. The effect size was derived from the findings of a small pilot study ($N = 16$) using the same design. The power analysis determined the required sample size to be 51 participants. This estimate was then increased to 60 subjects to account for: (1) the possibility that a percentage of participants would perform at ceiling on the Free-Response tests and thus be excluded from the key partial learning analyses of interest; (2) the fact that power analyses based on pilot data tend to underestimate required sample sizes (Albers & Lakens, 2018); and (3) to allow for equal number of participants across experimental lists (see *Procedure* below).

Participants were 60 adults (28 female, 27 male, five other/unknown; $M_{\text{age}} = 28$ years, $SD_{\text{age}} = 3.86$) recruited from the United States and the United Kingdom through the online subject pool Prolific (www.prolific.co). Participants self-identified as (multiple selections permitted): Asian (7), Black (4), White (46), multiracial (3), or other (2); three participants identified as Hispanic/Latinx. Although all instructions were in written English, participants were not required to be native English speakers (19 reported speaking another language in addition to English). Participants completed the study online via the Gorilla platform (Anwyl-Irvine et al., 2020). Participants were required to use a desktop computer or laptop and not a mobile device to participate. Participants received $5 for completing the study. All experimental procedures were approved by the University of Connecticut Institutional Review Board.

2.1.2. Stimuli

2.1.2.1. Target nouns and vignettes: The stimuli for this study were picture vignettes of eight target nouns known to be acquired early by English-learning children (Frank, Braginsky, Yurovsky, & Marchman, 2017): "apple," "bird," "book," "dog," "door," "flower," "hat," and "shoe." For each target noun, 12 vignettes were created from scenes in published children's picture books that had contained that noun in their original text. Three criteria guided the selection of the target vignettes. First, each vignette for a given target noun came from a unique book (i.e., each book contributed a maximum of one vignette per target noun[1]). Second, each vignette had a depiction of the referent somewhere in the scene (i.e., no instances of absent reference). Finally, all vignettes had been determined to be high in their referential ambiguity via prior norming studies (see *Target Vignette Norming* below). Each vignette was created by scanning picture book pages and then editing them so as to remove all of the text on the page (see Fig. 1a). The picture books used as source material for the vignettes included a mixture of classic and current picture books geared toward young children (book and stimulus details are described in more detail at the OSF site associated with this project: https://osf.io/xgwza/).

2.1.2.2. Target vignette norming: Because the focus of this work was to examine learning from referentially ambiguous naming events, a critical first step was to establish the ambiguity level of the target vignettes. Thus, all target vignettes were included in a series of norming studies designed closely after previous experiments assessing the referential ambiguity of naming events in child-directed speech (see *Experiment 1* in Medina et al., 2011; *Experiment 1* in Yurovsky et al., 2013). In our norming studies, participants were shown between 20 and 40 noun vignettes one at a time, with each vignette coming from a unique picture book and belonging to a different noun (i.e., a different noun was the correct answer for each vignette; more nouns than the current eight target nouns were part of the norming studies). For each vignette, participants were asked to type in a noun they thought "fits best on this page" (see Fig. 1b). Each vignette was rated by 16 different norming study participants. Vignette referential ambiguity was measured as the proportion of participants who guessed the correct noun. As in Medina et al. (2011; see also Gillette et al., 1999), guesses were coded as correct if they were an exact match with the correct noun or if they shared the same root morpheme as the correct noun but differed in grammatical number (e.g., "books" for "book") or included a diminutive suffix (e.g., "doggy" for "dog"). Following Medina and colleagues, vignettes for which fewer than 0.33 participants guessed the correct noun were classified as "referentially ambiguous" (Medina and colleagues used the term "low informative"). The mean ambiguity of the target vignettes used in this study was 0.04 ($SD = 0.08$, $Range = 0$–0.31).

2.1.2.3. Distractor vignettes: For the purposes of the 2AFC test (see *Experimental Design*), distractor vignettes were created to be paired with each target vignette. Distractor vignettes were randomly selected scenes from children's picture books that matched the target age range (within 0.5 years of the median) of those from which its target vignette pair was selected. Although distractor vignettes were not permitted to contain the target noun in their original *text*, distractor vignettes were permitted to contain the target noun's referent in

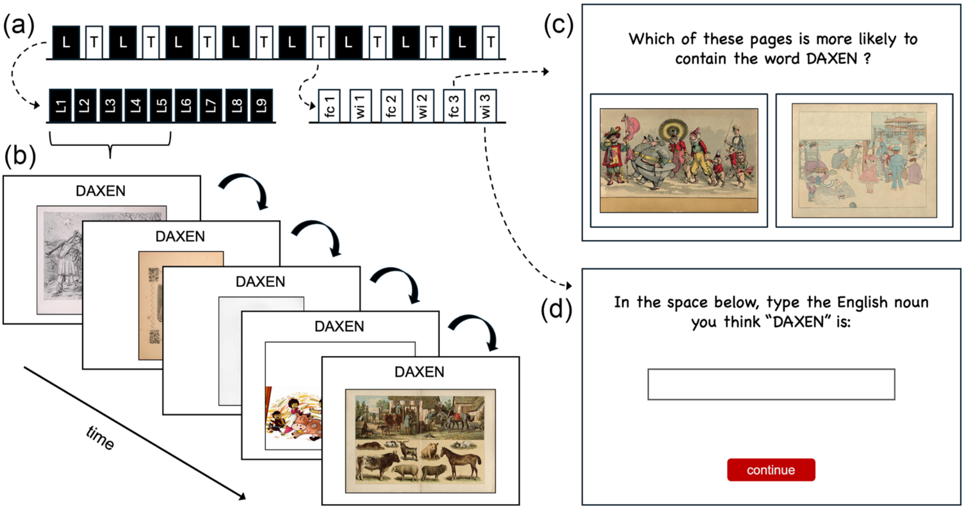

Fig. 2. Schematic of experimental design and examples of trials.
Note. This figure depicts the structure of the experimental design (Panel a; *L*: learning phase*;* *T*: test phase*;* *fc*: 2-alternative forced choice trial; *wi*: Word Identity trial), a sequence of five learning phase trials (Panel b), a sample 2-Alternative forced choice test trial (Panel c), and a sample Word Identity test trial (Panel d). Due to copyright issues, we are unable to depict the actual stimuli used in the current study. These scenes were selected to depict the ambiguity and visual complexity levels of the vignettes used in the current study. These scenes come from books either in the public domain or released under CC BY 4.0 license (see Harris, 1862; Caldecott, Evans, & Dobson, 1887; Frederick Warne, n.d., Hopkins, 1919; Mcloughlin Bro's, 1888; Nayar & Verulkar, 2018).

the *scene*.[2] Distractor vignettes were created using the same procedures described above for creating the target vignettes.

2.1.3. Experimental design

The experimental design was a blocked design with eight blocks, one per target noun (see Fig. 2a). Each target noun was assigned a disyllabic novel word that followed English phonotactic rules ("bemkin," "corwit," "daxen," "foppick," "hiplex," "renkle," "seebow," "tanzer"). Each block consisted of a learning phase followed by a testing phase. During the learning phase, participants were shown nine target vignettes one at a time. Each target vignette appeared centered on the screen with its novel word printed above it (Fig. 2b). The vignette and word remained on the screen for 3000 ms before the message "Press spacebar to continue" appeared below the vignette. The learning phase was self-paced, and the vignettes remained on the screen until the participants pressed the spacebar. The test phase immediately followed the nine learning trials and consisted of three pairs of trials. Each trial pair included two tests. The first test was a 2AFC trial in which participants were presented with two vignettes: a new target vignette (one they had not seen during the learning phase) and a distractor vignette placed side by side (see Fig. 2c). Participants' task in the 2AFC trial was to choose which vignette was more likely to contain the novel word. The second test in each trial pair immediately followed the 2AFC test and was a Word Identity (WI) free response test in which participants were asked to guess the English noun that the novel word represented (see Fig. 2d).[3] The 2AFC-WI trial structure was carried out three times in total per target noun, each with a new target and distractor vignette that had not been seen during the learning phase or in previous test trials. Participants were not given any feedback on any of the 2AFC or WI test trials. Because participants completed eight blocks (one per target noun), each participant contributed a total of 24 2AFC and 24 WI test trials. Thus, participants completed the full sequence of three 2AFC-WI trial pairs regardless of whether they guessed correctly in any of the WI trials in the sequence.

2.1.4. Procedure

Prior to the experiment, participants provided consent and completed a brief demographics questionnaire. Participants were then informed that they would see pages from children's picture books with their text removed, paired with a novel word representing a noun that had originally appeared in the text. Participants were told that their task was to figure out the English noun represented by the novel word. To familiarize participants to the task, they first completed three practice blocks consisting of four learning trials and three test trial pairs. The practice blocks used novel words ("habble," "modi," "tema") and target vignettes belonging to English nouns ("cow," "car," and "cake") that were not used in the experiment proper. After each practice block, participants were shown the correct answers for each 2AFC and WI practice test trial. The first experiment block started immediately after the three practice sets. Unlike the practice blocks, participants were not told the correct answers following the experiment blocks.

Participants were randomly assigned to one of four experimental lists that differed only in which vignettes served as learning phase vignettes versus testing phase vignettes. Across the four lists, each target vignette (and its corresponding distractor vignette) appeared once as part of the 2AFC test. Participants took 15–30 min to complete the entire study and were asked not to take breaks or take any written notes during the experiment. The project's OSF site (https://osf.io/xgwza/) hosts a screencast depicting the study procedure.

2.2. Results

2.2.1. Data analysis approach

Our data analyses are presented as follows. We first present data on participants' overall performance in the 2AFC and WI test trials. We then present the key partial learning analyses of interest, which are participants' performance in the 2AFC test trials when they guessed the target noun incorrectly in the corresponding WI test trial. Of particular interest was whether participants would select the correct scene on the 2AFC test trials that preceded an incorrect WI guess, which would indicate that they had some *partial* knowledge of the word's meaning despite failing to identify it exactly. For example, in cases where the target noun had been "apple," of interest was whether participants would select the 2AFC scene in which "apple" had appeared in the original text even when they guessed the target noun was something other than "apple" in the WI test (e.g., "food," "tree," etc.). Because our primary analyses of interest

Table 1
Descriptive statistics for two-alternative forced choice (2AFC) and Word Identity (WI) performance in Experiment 1

Test Trial Type	All Trials	Trial Order (Within Target Noun)		
		Trial 1	Trial 2	Trial 3
2AFC test	0.80 (0.09)	0.81 (0.14)	0.80 (0.14)	0.78 (0.14)
WI test	0.56 (0.24)	0.58 (0.25)	0.55 (0.24)	0.56 (0.25)

Note. This table reports the means and standard deviations (in parentheses) of participants' 2AFC and WI test trials in Experiment 1.

focused on an unpredictable subset of the data (i.e., 2AFC trials where participants guessed incorrectly on the corresponding WI trial), the number of trials contributed by each participant and by each target noun to this specific analysis was unequal and, in some cases, small. We suspect that for this reason, our attempt to utilize mixed-effects modeling on the current data (Baayen, Davidson, & Bates, 2008; Jaeger 2008) frequently resulted in models with singular fits (see Bolker, 2023). Thus, we opted for a more traditional statistical approach to the current dataset - t-tests and analysis of variance (ANOVA) over participant-level proportions. We do provide documentation on the mixed-effects modeling of these data in the Supporting Information. It is worth noting that all key partial learning results reported here with more traditional statistics are confirmed by the mixed effects models. To account for potential non-normality in the data, nonparametric bootstrap resampling (with the bias-corrected and accelerated [BCa] method, Efron and Tibshirani, 1994) was employed to generate 95% confidence intervals (CI) around the key means of interest (5000 bootstrap resamples implemented with the "boot" package in R version 4.5.1; Canty & Ripley, 2024).

2.2.2. *Overall performance*

2.2.2.1. 2AFC test performance: For each participant, we computed the proportion of 2AFC test trials in which they selected the correct answer. As a group, participants were highly accurate in the 2AFC test, guessing correctly on 0.80 of trials ($SD = 0.09$), a proportion that is well above the rate of guessing, $t(59) = 24.86$, $p < .001$, $d = 3.21$. We then explored whether participants' 2AFC test performance improved over the course of the three 2AFC test trials. As displayed in Table 1, participants' 2AFC test performance was more or less stable. Indeed, a one-way repeated-measures ANOVA revealed that 2AFC performance across the three trials did not differ from one another, $F(2,118) = 0.88$, $p = .417$.

2.2.2.2. WI test performance: Using the same guidelines for coding guesses in the norming studies (see *Target Vignette Norming* above; see also Gillette et al., 1999; Medina et al., 2011), participants' response on each WI test trial was scored as correct or incorrect.[4] Participants' mean proportion correct on all WI test trials was 0.56 ($SD = 0.24$). Unexpectedly, a one-way repeated-measures ANOVA revealed that performance across the three WI test trials was unequal, $F(1.86,118) = 4.92$, $p = .010$[5] (see Table 1). Pairwise comparisons using Bonferroni corrections indicated that participants' were more accurate in their guesses on the first

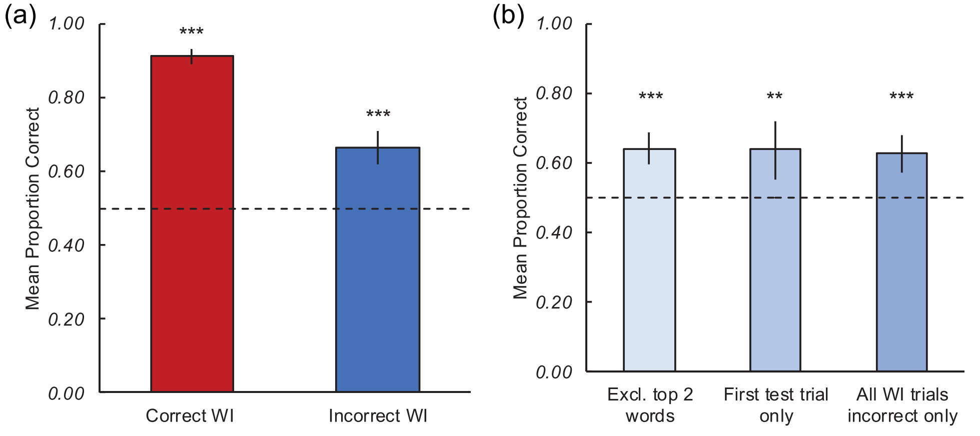

Fig. 3. Two-alternative forced choice (2AFC) performance as a function of Word Identity (WI) accuracy in Experiment 1.
Note. Panel (a) depicts 2AFC performance on trials prior to a correct and incorrect WI guess. Panel (b) depicts three validation tests of the partial learning effects: (left) mean 2AFC performance prior to an incorrect WI guess excluding the two words for which participants exhibited the greatest partial learning, (center) mean 2AFC performance prior to an incorrect WI guess restricted to only the first test trial, (right) mean 2AFC performance prior to an incorrect WI guess restricted to only words for which participants never offered a correct WI guess.
** $p < .01$; *** $p < .001$. Error Bars represent 95% bias-corrected and accelerated (BCa) bootstrapped CIs around means.

WI test trial than on the second, $t(59) = 2.82$, $p < .05$, $d = 0.36$; there were no differences between the first WI and third WI test trials, $p = .11$ or between the second and third WI test trials, $p = .60$.

2.2.3. *2AFC performance when WI guess was incorrect*

Of most interest in the current study was participants' performance on the 2AFC test trials when they had incorrectly identified the target noun in the corresponding WI test trial. Thus, we next examined participants' 2AFC performance as a function of their WI performance. Unsurprisingly, and as depicted in Fig. 3a, participants were more likely to select the correct vignette in 2AFC test trials prior to a correct WI guess ($M = 0.91$, $SD = 0.08$, 95% BCa bootstrap confidence interval [0.89–0.93]) than prior to an incorrect WI guess ($M = 0.66$, $SD = 0.18$, 95% BCa CI [0.62–0.71]), $t(58) = 9.41$, $p < .001$, $d = 1.22$.[6] Importantly, however, participants' 2AFC performance prior to an incorrect WI guess was still well above chance levels, $t(58) = 7.15$, $p < .001$, $d = 0.93$. This finding suggests that when participants failed to acquire the precise meanings of words from referentially ambiguous scenes, they nonetheless extracted partial information that allowed them to succeed reliably in the 2AFC word learning test.

2.2.3.1. Validation tests: Three follow-up tests were conducted to validate these partial knowledge effects (see Fig. 3b). First, we asked whether these findings were driven simply by performance in just one or two target nouns. Thus, we recomputed participants' 2AFC performance prior to an incorrect WI test trial but now excluded their performance on the two words (*apple* and *book*) for which participants, as a group, exhibited the most partial learning (see the Appendix for detailed by-word trends). Participants' recomputed 2AFC performance prior to an incorrect WI guess ($M = 0.64$, $SD = 0.18$, 95% BCa CI [0.59–0.69]) was still reliably above chance levels, $t(58) = 5.86$, $p < .001$, $d = 0.76$.[7] Second, we considered the possibility that the partial learning effects may have been driven by learning during the three test trials, as opposed to the product of learning from the learning phase. Thus, a second follow-up analysis was conducted on only the first AFC-WI test pair in each block and revealed that participants' 2AFC performance prior to an incorrect initial WI guess ($M = 0.64$, $SD = 0.34$, 95% BCa CI [0.55–0.72]) was reliably above chance levels, $t(57) = 3.19$, $p < .01$, $d = 0.42$. Finally, we examined whether the observed partial learning effects were driven by a subset of participants who had entertained the correct target noun in at least one of the three WI trials. Although these participants would have nonetheless exhibited a dissociation between AFC and WI performance (on the trials where they were incorrect in their WI guess), these participants could be viewed as having been very close to precise word meaning identification. Thus, we subsetted our AFC dataset to include only blocks where participants were incorrect in all three of their WI test trials (573 AFC trials over 58 participants; 92% of total AFC trials that preceded an incorrect WI guess). In this reduced dataset (see Fig. 3b), participants' 2AFC performance ($M = 0.63$, $SD = 0.21$, 95% BCa CI [0.57–0.68]) was still reliably above chance levels $t(57) = 4.50$, $p < .001$, $d = 0.59$ validating the observed dissociability between successful 2AFC selections and successful WI guesses.

2.3. Discussion

The findings from Experiment 1 show that when participants fail to identify word meanings from referentially ambiguous events in a free response task, they nonetheless demonstrate some word knowledge in an alternative-forced choice task. Of course, dissociations across tasks are nothing new to the word learning field (e.g., Gershkoff-Stowe & Hahn, 2013; Gurteen et al., 2011; Hendrickson et al., 2015; Horst & Samuelson, 2008). Nevertheless, given the ongoing debates over the kinds of input that matter for word learning (Cartmill et al., 2013; Medina et al., 2011; Yurovsky et al., 2013; Zhang et al., 2021), as well as the implications of that input for debates on the mechanisms of learning (Gleitman & Trueswell, 2020; Yu, Zhang, Slone, & Smith, 2021), the current findings are a reminder that the resolution of these debates may depend in part on how the output of learning is measured (see also Smith et al., 2014; Yu & Smith, 2012a). The current findings are also a reminder that word meaning knowledge falls on a continuum and that acquiring "full knowledge" is a slow and protracted process (Carey, 2010; Carey & Bartlett, 1978; Clark, 2023; Kucker, McMurray, & Samuelson, 2015; see also Ameel et al., 2008; Deák & Toney, 2013; Hendrickson et al., 2015; Lo et al., 2021). Yet, as many have demonstrated, word learners do not await full word knowledge before their representations organize attention and memory (Baldwin & Markman, 1989; LaTourrette & Waxman, 2020; Twomey & Westermann, 2018), support new word learning (e.g., Fitneva & Christiansen, 2011; Yurovsky et al., 2014), and shape their communicative acts with those words (Clark, 2023; Shatz, Tare, Nguyen, & Young, 2010). Thus, if referentially ambiguous events support partial word learning, even when they fall short in supporting precise meaning acquisition, then referentially ambiguous events may nonetheless confer many benefits to cognitive, linguistic, and socio-communicative development.

One limitation and reasonable concern about the current findings is participants' high overall success rate in the word identity test. That is, participants in the current study accurately identified the exact noun in over 50% of trials. This rate of learning from referentially ambiguous events is greater than the rate observed by Medina et al. (2011), who reported exact noun identification from referentially ambiguous events in approximately only 20% of WI trials (see the "HI Absent" condition in Medina et al.'s Experiment 2). There are many candidate explanations for this difference, including the nature of the events (picture-book vs. video-based vignettes), task differences (e.g., massed vs. interleaved presentation), and the precise nouns that were tested. Regardless of the explanation, however, the high rate of word identity guesses raises the possibility that the observed partial word learning in the current study is limited to referentially ambiguous events that are capable of eliciting a considerable degree of exact meaning identification. The goal of Experiment 2 was to test this possibility.

3. Experiment 2

In Experiment 2, we asked whether learners could still extract partial word knowledge from referentially ambiguous events if those events were highly referentially ambiguous. Nearly all aspects of the experimental design and procedure of Experiment 2 were the same as those of Experiment 1. However, all of the target vignettes in Experiment 2 had maximal referential ambiguity, which was operationalized as zero correct word identifications in the vignette norming process. We hypothesized that restricting the vignettes in Experiment 2 to those with maximal ambiguity would lead to a significant reduction in learners' success in the word identity test. Of most interest, however, is whether despite the reduction in exact word meaning identification, learners would nevertheless still exhibit partial knowledge of those word meanings.

3.1. Methods

3.1.1. Participants

The sample size for Experiment 2 was determined through the same procedure as for Experiment 1. An unexpectedly large effect size in the pilot for Experiment 2, however, suggested that the estimated sample size needed for Experiment 2 was smaller than the one for Experiment 1. Due to the distinct possibility that effect size estimates from pilot data may lead to underestimates of sample size requirements (Albers & Lakens, 2018), we decided to collect the same sample size for Experiment 2 as was collected for Experiment 1. Thus, participants in Experiment 2 were 60 adults (31 female, 26 male, three other/unknown; $M_{\text{age}} = 28$ years, $SD_{\text{age}} = 4.78$) who had not participated in Experiment 1. Participants were recruited

Table 2
Descriptive statistics for 2AFC and WI performance in Experiment 2

Test Trial Type	All Trials	Trial Order (Within Target Noun)		
		Trial 1	Trial 2	Trial 3
2AFC test	0.73 (0.12)	0.75 (0.17)	0.72 (0.18)	0.71 (0.17)
WI test	0.29 (0.26)	0.29 (0.26)	0.30 (0.26)	0.29 (0.27)

Note. This table reports the means and standard deviations (in parentheses) of participants' 2AFC and WI test trials in Experiment 2.

from the United States and the United Kingdom through Prolific and completed the study on Gorilla via a desktop or laptop computer. Fifty-seven of the participants provided racial and ethnic demographic information and self-identified as Black (3), White (49), multiracial (5), or other (1); one participant identified as Hispanic/Latinx. Eighteen participants reported speaking another language in addition to English. Participants received $5 for completing the study.

3.1.2. Stimuli, experimental design, and procedure

The experimental design, procedures, and data coding for Experiment 2 were identical to those for Experiment 1. The key methodological difference between Experiments 1 and 2 was the ambiguity level of the target vignettes. Recall in Experiment 1 vignettes were under consideration if their ambiguity level, as determined via norming studies, was below 0.33 (see *Target Vignette Norming* above for norming study details). In Experiment 2, all target vignettes had the ambiguity level of 0. This means that for all Experiment 2 target vignettes, none of the norming study participants had correctly identified the target noun from the scene. Sixty-six of the 96 target vignettes used in Experiment 1 fit this criterion and were used again in Experiment 2. The 30 target vignettes in Experiment 1 that did not fit this criterion were replaced with new target vignettes that did,[8,9] For each of the 30 new target vignettes, a new distractor vignette was identified and created following the protocol described in Experiment 1.[10]

3.2. Results and discussion

3.2.1. Overall performance

3.2.1.1. 2AFC test performance: Participants in Experiment 2, as a group, performed well in the 2AFC test, guessing correctly at a rate that is well above chance levels (0.50), $M = 0.73$, $SD = 0.12$, $t(59) = 15.04$, $p < .001$, $d = 1.94$. Overall 2AFC performance in Experiment 2 was, however, reliably lower than overall 2AFC performance in Experiment 1 ($M = 0.80$; $SD = 0.09$), $t(118) = -3.74$, $p < .001$, $d = 0.68$, highlighting that the reduced informativity of target vignettes in Experiment 2 made for a more challenging learning task. 2AFC test performance did not differ across the three 2AFC test trials per word (see Table 2), $F(2,118) = 0.89$, $p = .414$.

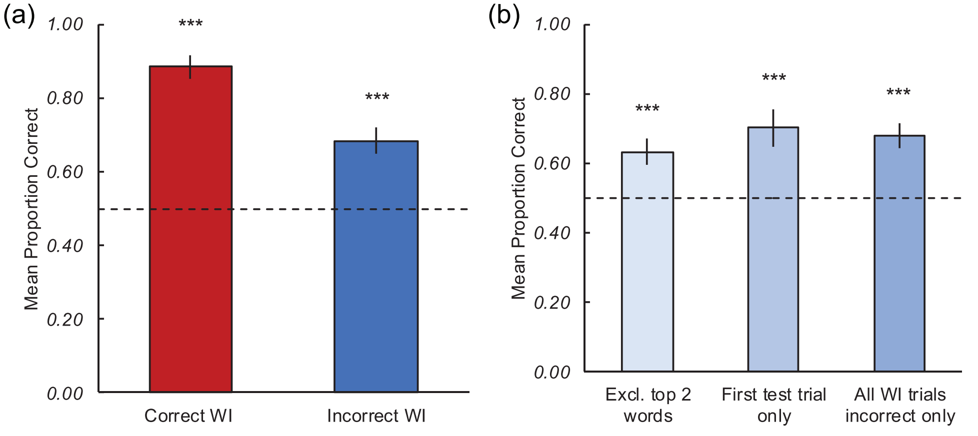

Fig. 4. 2AFC performance as a function of WI accuracy in Experiment 2.
Notes. Panel (a) depicts 2AFC performance on trials prior to a correct and incorrect WI guess. Panel (b) depicts three validation tests of the partial learning effect: (left) mean 2AFC performance prior to an incorrect WI guess excluding the two words for which participants exhibited the greatest partial learning, (center) mean 2AFC performance prior to an incorrect WI guess restricted to only the first test trial, (right) mean 2AFC performance prior to an incorrect WI guess restricted to only words for which participants never offered a correct WI guess.
Note: *** $p < .001$. Error Bars represent 95% BCa bootstrapped CIs around means.

3.2.1.2. WI test performance: Participants' mean proportion correct on all WI test trials was 0.29 ($SD = 0.26$).[11] WI performance in Experiment 2 was lower than that in Experiment 1 ($M = 0.56$; $SD = 0.24$), $t(118) = -5.86$, $p < .001$, $d = 1.07$, illustrating the negative effect that reducing target vignette informativity had on precise word meaning acquisition. As displayed in Table 2, participants performance across the three WI test trials was stable, $F(1.67,118) = 0.41$, $p = .635$.[12]

3.2.2. 2AFC performance when WI guess was incorrect

As in Experiment 1, the main analysis of interest in Experiment 2 was participants' performance on the 2AFC test trials when they had incorrectly identified the target noun in the corresponding WI test trial. As illustrated by Fig. 4a, participants were more likely to select the correct vignette in the 2AFC test trial prior to a correct WI guess ($M = 0.89$, $SD = 0.11$, 95% BCa CI [0.85–0.91]) than prior to an incorrect WI guess ($M = 0.68$, $SD = 0.14$, 95% BCa CI [0.65–0.72]), $t(45) = 6.93$, $p < .001$, $d = 1.02$.[13] Importantly, and as was the case in Experiment 1, participants' 2AFC performance prior to an incorrect WI guess was nevertheless above chance levels, $t(59) = 10.20$, $p < .001$, $d = 1.32$. Interestingly, although, as reported above, there were differences in overall 2AFC and WI performance between Experiments 1 and 2, there were no differences between the two experiments in partial learning (i.e., rate of correct AFC trials prior to incorrect WI trials; $M_{Exp\ 1} = 0.66$; $M_{Exp.\ 2} = 0.68$), $t(117) = 0.67$, $p = .506$, $d = 0.12$. As with Experiment 1, we conducted three validation tests of Experiment 2's partial learning effects (see Fig. 4b). First, in an analysis that excluded performance for the two words on which participants as a group exhibited the most partial learning (*apple* and *bird*; see Table A2 in the Appendix), we found that participants' 2AFC performance prior to an incorrect WI guess ($M = 0.63$, $SD = 0.15$, 95% BCa CI [0.59–0.67]) was still reliably above chance levels, $t(58) = 6.68$, $p < .001$, $d = 0.87$. Second, these partial learning effects held when partial learning was analyzed for only the first 2AFC-WI test pair in a given block ($M = 0.70$, $SD = 0.22$, 95% BCa CI [0.65–0.76]), $t(59) = 7.01$, $p < .001$, $d = 0.90$. Finally, these partial learning effects also held when we restricted our analyses to only words for which participants never guessed the correct target noun on any of the three WI test trials ($M = 0.68$, $SD = 0.15$, 95% BCa CI [0.64–0.71]), $t(59) = 9.49$, $p < .001$, $d = 1.22$.

3.2.3. Discussion

Experiment 2 was designed as a test of the reliability of the partial learning effects observed in Experiment 1. The increase in referential ambiguity of the target vignettes in Experiment 2 had the anticipated effect of reducing overall amounts of word learning as measured by a word identity free response assessment. Despite this reduction in participants' ability to identify precise word meanings, participants' ability to extract partial word knowledge was unchanged. That is, even though participants in Experiment 2 struggled to identify the correct target noun from the highly referentially ambiguous scenes, they nevertheless were able to distinguish scenes that did and did not co-occur with those target nouns. These data reinforce the conclusion from Experiment 1 that although referentially ambiguous events may not afford precise meaning acquisition, they do allow learners to extract valuable partial word knowledge that they can then successfully deploy in some word learning tasks.

4. General discussion

Understanding the kinds of input that shape children's word learning has broad implications, from a better understanding of the mechanisms that shape learning (see Gleitman & Trueswell, 2020; Yu et al., 2021) to informing best practices for input-based interventions (see Masek, Ramirez, McMillan, Hirsh-Pasek, & Golinkoff, 2021; Rowe & Snow, 2020). Previous research on whether word learners can leverage referentially ambiguous naming events for learning has yielded mixed results. Many studies have shown robust learning from referentially ambiguous events when those events consist of a few novel objects (for review, see Roembke et al., 2023). Other studies have shown very little learning from referentially ambiguous events when those events capture real-world levels of ambiguity (see LaTourette et al., 2022; Medina et al., 2011). Finally, some research suggests that whether one can learn words from real-world levels of referential ambiguity depends on the perspective from which those events are viewed (i.e., through a first-person learner's viewpoint vs. a third-person observer's viewpoint; Yurovsky et al., 2013; Zhang et al., 2021). The current study sheds new light on the potential for referentially ambiguous naming events to move the needle of learning. In a paradigm that assessed word learning in multiple ways, the current study reveals that when referentially ambiguous events fail to yield learning of *exact* word meanings, they nevertheless yield partial learning of word meanings. In other words, referentially ambiguous naming events may allow learners to find the right region of semantic space even when they do not reveal the exact coordinates within that region.

4.1. Partial word learning in early vocabulary development

Some may view that the key to understanding early vocabulary development is understanding the acquisition of precise word meanings or the processes that would allow for success in assessments that resemble the word identification test of the current study. Under this view, the current partial learning effects of success in an alternative forced choice test coupled with failure in a word identification test might be viewed as a negligible part of word learning. At least two datapoints however suggest the value of taking more seriously partial learning effects like those observed in the current study. First, many of the methodologies used in clinical, educational, and research settings to assess vocabulary development (e.g., infant looking while listening paradigms Fernald, Zangl, Portillo, & Marchman, 2008; Peabody Picture Vocabulary Test, Dunn, 2019) are actually closer in format to alternative forced choice tests than they are to word identification tests. Thus, partial word knowledge may often be what children possess when it is said that children know the meaning of a word. Second, there are many empirical demonstrations that children go through protracted stages of learning in which they possess some sense of a word's meaning while falling short of the adult sense of that word's meaning. Demonstrations of this type of partial knowledge date back at least to Carey and Bartlett's classic experiments on color-term acquisition (Carey & Bartlett, 1978; see also Wagner et al., 2013) but have also been reported in children's acquisition of object names (e.g., Ameel et al., 2008; Hendrickson et al., 2015; Lo et al., 2021), action words (e.g., Saji et al., 2011), time words (e.g., Shatz et al., 2010; Tillman & Barner 2015), number words (e.g., O'Rear et al., 2020; Wagner, Chu, & Barner, 2019), and emotion terms (e.g., Hoemann, Xu, & Barrett, 2019; Widen & Russell, 2008), to name a handful of examples. Thus, we suggest that there is more to be gained from a broader conception of word learning that incorporates partial learning rather than a narrower one that ignores it (see also Wojcik, Zettersten, & Benitez, 2022).

The key contribution of the current study is of course not that partial word learning occurs but rather the potential for referentially ambiguous naming events to contribute to such learning. To the best of our knowledge, the vast majority of research on partial word knowledge focuses more on documenting the presence of that partial word knowledge rather than understanding the kinds of input that could contribute to it. In considering the handful of studies that have investigated the input to partial learning (Dautriche & Chemla, 2014; Roembke & McMurray, 2016; Yurovsky & Frank, 2015; Yurovsky et al., 2014), the current work is unique in the nature and degree of the referentially ambiguous events presented to learners. Unlike previous studies whose ambiguous scenes consist of a few novel objects depicted on a screen, the current referentially ambiguous events were complex, multifaceted scenes whose ambiguity levels were determined using an established methodology (see Cartmill et al., 2013; Medina et al., 2011). Although the current study's referentially ambiguous naming events were not taken from recordings of child-directed speech (see Medina et al., 2011; Yurovsky et al., 2013; Zhang et al., 2021), it is important to note that naming events in picture books do represent an important real-world source of input for many children the world over (see Horst & Houston-Price, 2015; Montag et al., 2015; Nation et al., 2022).

4.2. Speculations on the nature of partial learning via referentially ambiguous events

Although the current work suggests that partial word knowledge could emerge from referentially ambiguous naming events, it does not reveal the details of this process. One open question pertains to the information that learners extracted from these scenes that allowed them to succeed in the alternative forced choice test without succeeding in the word identity test. One possibility is that learners honed in on specific competing semantically related referents and mistakenly mapped the novel word to those referents. For example, participants may have mismapped the novel word that corresponded to the English word "flower" to a meaning like *plant* or *leaf*. Although such mismappings would not have allowed participants to pass the stringent criteria of the word identity test, they would have likely allowed participants to perform well in the forced choice test since the target vignette would, on average, be more likely to also depict those closely related referents.

A separate possibility is that participants succeeded in the forced choice test not by mapping novel words to any specific closely related meanings but by extracting a broader "gist" of the referentially ambiguous scenes (e.g., "school scene," "park scene," "house scene," etc.; see Oliva, 2005). We suggest this as a possible explanation because the different-to-be-learned words did tend to occur in different situational contexts, akin to how different words in child-directed speech have been reported to occur (see Roy, Frank, DeCamp, Miller, & Roy, 2015; Tamis-LeMonda, Custode, Kuchirko, Escobar, & Lo, 2019). For example, the referentially ambiguous scenes for "bird" and "flower" were more likely to be scenes of the outdoors, the scenes for "book" were more likely to depict indoor settings, and the scenes for "hat" and "shoe" were more likely to be scenes containing people. Thus, perhaps it is these broad gists shared across a word's referentially ambiguous scenes that allowed participants to identify new scenes most likely belonging to the to-be-learned word. Future research that incorporates additional measures into the current paradigm (e.g., eye tracking, post-experiment interviews) may help clarify the nature of the partial knowledge that supported learners' success, the specific information they relied on to extract that knowledge, and the extent to which they could articulate what they had learned.

A second and related open question pertains to the nature of the mechanisms that allowed for the observed partial learning. One potential mechanism is based on the idea of word learning as a hypothesis testing process (see Aravind et al., 2018; Stevens, Gleitman, Trueswell, & Yang, 2017; Trueswell, Medina, Hafri, & Gleitman, 2013). Under this account, the observed partial learning effects (i.e., success in the forced choice test in the absence of success in the word identity test) would be the product of participants maintaining a single hypothesis that, albeit incorrect, was nevertheless close enough in meaning to afford success on the forced choice test trials. For example, the incorrect hypotheses "fruit" or "tree" for the novel word that meant "apple" would likely yield above chance performance in the forced choice test because fruits and trees, as well as things related to them, are more likely depicted in the correct alternative forced choice scene. An alternative mechanism to explain these data is based on associative models of word learning (see Kachergis & Yu, 2018; McMurray, Horst, & Samuelson, 2012; Yu & Smith, 2012a). Under this account, learners extract and maintain in memory much more than a single hypothesized meaning, including memories of various objects, actions, and features depicted in these scenes (Chen & Yu, 2017; Dautriche & Chemla, 2014; Knabe & Vlach, 2023; Zettersten, Wojcik, Benitez, & Saffran, 2018). In such a model, the partial learning effects may be the product not only of a close-but-incorrect hypothesis but also of memory traces of the events peripheral to any hypotheses.

As is now well documented (see Bhat, Spencer, & Samuelson, 2022; Yu & Smith, 2012a), distinguishing between these two classes of mechanisms is no easy task, and by no means was the current work designed to aid in distinguishing between them. It does however offer one lesson relevant to this thorny issue. That is, many current discussions on the nature of word learning mechanisms are often merged with discussions on the kinds of input that contribute to learning. For example, some proponents of hypothesis testing models argue for a learning process that is primarily driven by the referentially transparent naming events in the input (or the “gems” in the so called “gems and junk” model of learning; see Gleitman & Trueswell, 2020; Medina et al., 2011). On the other hand, proponents of the statistical associative learning models tend to argue for a learning process that is driven by a much wider range of events, including the more referentially ambiguous one (see Yu & Smith, 2012a; Zhang et al., 2021). Considering that the current work reveals the contributions of referentially ambiguous naming events for learning without distinguishing between the two models, we suggest that there may be value in treating the issue of whether referentially ambiguous events contribute to learning as connected but nevertheless separate from the issue of the mechanisms that underlie such learning.

4.3. Limitations and future directions

There are several limitations to the current studies that are worth noting and worth exploring in future work. First, although care was taken to ensure that the level of referential ambiguity in the current picture book naming events was comparable to that in previous studies investigating learning via child-directed naming events (e.g., Medina et al., 2011), it is possible that the ambiguity in children’s picture books is more amenable to partial learning than the ambiguity in child-directed speech. Thus, it would be important to test the generalizability of the current findings to referentially ambiguous events in child-directed speech. Second, there are features of the current study’s to-be-learned words (e.g., all were object names), naming events (e.g., none depicted cases of absent reference), and experimental paradigm (e.g., vignettes followed a massed presentation schedule) that are likely to have increased the likelihood of partial learning. Although the exact degree to which some of these features are a departure from children’s real-world learning input is a continued matter of investigation (e.g., see Slone, Abney, Smith, & Yu, 2023), these methodological decisions nevertheless suggest the need to push the limits of partial learning via referentially ambiguous events. Finally, like many cross-situational word learning (e.g., Yu & Smith, 2007) and Human Simulation Paradigm studies (e.g., Gillette et al., 1999), the current work used adult learning as a model for children's learning. Although some have revealed similarities in how adults and children learn in these paradigms (e.g., Piccin & Waxman, 2007; Suanda, Mugwanya, & Namy, 2014), others have revealed differences (e.g., Benitez, Zettersten, & Wojcik, 2020; Fitneva & Christiansen, 2017). Thus, it would be critical to ask whether, like the adults in the current study, young children can extract and deploy partial knowledge from complex referentially ambiguous events.

4.4. Conclusion

The role that referentially ambiguous naming events play in children's word learning has long stimulated interest and research into how children learn the meanings of words. The current study highlights that inferences about the role that such events play in learning depend, at least in part, on how word learning is assessed and defined. If one constrains the definition of word learning to identifying the exact meaning intended by a speaker, then referentially ambiguous events alone may fall short in playing a significant role in learning. In contrast, if one considers partial but imperfect knowledge of a word's meaning as an important part of the learning process, then the current work suggests that referentially ambiguous events may play a valuable role in learning. Thus, under this latter conception of learning, there may be much more to referentially ambiguous events than initially meets the eye.

Acknowledgments

We thank Peter Schoener and the members of the UConn Communication and Development Lab (especially Miranda Long, Amanda Mankovich, Beth Stoddard, Menghan Yang, and Kosta Boskovic) for their input on the project, and Gerry Altmann for helpful comments on this manuscript. This research was supported in part by the James S. McDonnell Foundation (JSMF 220020549) and the National Institutes of Health (R00-HD082358).

Conflict of Interest Statement

There are no conflicts of interest to declare.

Data Availability Statement

The data for this manuscript are available at: https://osf.io/xgwza/.

Notes

1 Eighteen books contributed vignettes to multiple target nouns (16 books contributed to two target nouns, and two books contributed to three target nouns); across all 96 target vignettes, a total of 76 different books were used.

2 14% of the distractor vignettes contained the target noun's referent in the scene. Excluding trials with these distractors had no impact on the level of statistical significance (i.e., "$p < .001$," "$p < .01$," etc.) of the analyses.
3 The decision to have the 2AFC trial precede the WI trial was to ensure that participants observed the same amount of information prior to responding to both trial prompts.
4 On three WI trials (0.21% of all trials), participants entered multiple answers, one of which was correct. Such answers were scored as partially correct (e.g., 0.5 if they provided two answers, one of which was correct).
5 Mauchly's Test of Sphericity was violated ($X^2 = 6.38$, $p < .05$); reported are the test statistics with the Huynh–Feldt correction.
6 One participant in Experiment 1 guessed correctly on all WI trials and thus did not contribute data to this analysis.
7 Because not all participants made incorrect WI guesses for each word, the number of participants contributing to AFC performance following an incorrect WI guess was small for some words (*n*'s ranged from 11 to 44 across words). Thus, we did not feel we had sufficient power to analyze and report by-word findings.
8 One vignette from Experiment 1 was replaced despite fitting Experiment 2's ambiguity level criterion. This was due to a clerical error in the norming documentation where its ambiguity level had been marked as greater than 0.
9 Across the 96 target vignettes used for Experiment 2, a total of 81 different books were used. As in Experiment 1, each book contributed a maximum of one vignette per target noun; 11 books contributed vignettes to two different target nouns, and two books contributed vignettes to three different target nouns;
10 18% of Experiment 2 distractor vignettes contained the target noun's referent in the scene. Excluding trials with these distractors had no impact on the level of statistical significance (i.e., "$p < .001$," "$p < .01$," etc.) of the analyses.
11 On two WI trials (0.14% of all trials), participants entered multiple answers, one of which was correct. Such answers were scored as partially correct.
12 Mauchly's Test of Sphericity was violated ($X^2 = 12.55$, $p < .01$); reported are the test statistics with the Huynh–Feldt correction.
13 Fourteen participants were excluded from this analysis because they never guessed the target noun correctly in the WI test and thus did not contribute any data to 2AFC performance prior to a correct WI guess.

References

Albers, C., & Lakens, D. (2018). When power analyses based on pilot data are biased: Inaccurate effect size estimators and follow-up bias. *Journal of Experimental Social Psychology*, *74*, 187–195. https://doi.org/10.1016/j.jesp.2017.09.004

Ameel, E., Malt, B., & Storms, G. (2008). Object naming and later lexical development: From baby bottle to beer bottle. *Journal of Memory and Language*, *58*(2), 262–285. https://doi.org/10.1016/j.jml.2007.01.006

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, *52*, 388–407. https://doi.org/10.3758/s13428-019-01237-x

Aravind, A., de Villiers, J., Pace, A., Valentine, H., Golinkoff, R., Hirsh-Pasek, K., Iglesias, A., & Wilson, M. S. (2018). Fast mapping word meanings across trials: Young children forget all but their first guess. *Cognition*, *177*, 177–188. https://doi.org/10.1016/j.cognition.2018.04.008

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Baldwin, D. A., & Markman, E. M. (1989). Establishing word-object relations: A first step. *Child Development*, *60*(2), 381–398. https://doi.org/10.2307/1130984

Benitez, V. L., Zettersten, M., & Wojcik, E. (2020). The temporal structure of naming events differentially affects children's and adults' cross-situational word learning. *Journal of Experimental Child Psychology*, *200*, 104961. https://doi.org/10.1016/j.jecp.2020.104961

Bhat, A. A., Spencer, J. P., & Samuelson, L. K. (2022). Word-Object Learning via Visual Exploration in Space (WOLVES): A neural process model of cross-situational word learning. *Psychological Review*, *129*(4), 640–695. https://doi.org/10.1037/rev0000313

Bloom, P. (2000). *How children learn the meanings of words*. Cambridge: MIT Press.

Bloom, P. (2001). Précis of How Children Learn the Meaning of Words. *Behavioral and Brain Sciences*, *24*(6), 1095–1134. https://doi.org/10.1017/S0140525X01000139

Bolker, B. (2023). GLMM FAQ. Available at: https://bbolker.github.io/mixedmodels-misc/glmmFAQ. June 1 2024

Caldecott, R., Evans, E., & Dobson, A. & George Routledge and Sons. (1887). *The complete collection of pictures & songs*. New York: G. Routledge and Sons. https://www.loc.gov/item/42048476/

Canty, A., & Ripley, B. D. (2024). boot: Bootstrap R (S-Plus) functions. R package version 1.3-31. https://doi.org/10.32614/CRAN.package.boot

Carey, S., & Bartlett, E. (1978). Acquiring a single new word. *Proceedings of the Stanford Child Language Conference*, *15*, 17–29.

Carey, S. (2010). Beyond fast mapping. *Language Learning and Development*, *6*(3), 184–205. https://doi.org/10.1080/15475441.2010.484379

Cartmill, E. A., Armstrong III, B. F., Gleitman, L. R., Goldin-Meadow, S., Medina, T. N., & Trueswell, J. C. (2013). Quality of early parent input predicts child vocabulary 3 years later. *Proceedings of the National Academy of Sciences*, *110*(28), 11278–11283. https://doi.org/10.1073/pnas.1309518110

Chen, C.-H., & Yu, C. (2017). Grounding statistical learning in context: The effects of learning and retrieval contexts on cross-situational word learning. *Psychonomic Bulletin & Review*, *24*(3), 920–926. https://doi.org/10.3758/s13423-016-1163-x

Clark, E. V. (2023). A gradualist view of word meaning in language acquisition and language use. *Journal of Linguistics*, *59*(4), 737–762. https://doi.org/10.1017/S0022226722000330

Dautriche, I., & Chemla, E. (2014). Cross-situational word learning in the right situations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(3), 892–903. https://doi.org/10.1037/a0035657

Deák, G. O., & Toney, A. J. (2013). Young children's fast mapping and generalization of words, facts, and pictograms. *Journal of Experimental Child Psychology*, *115*(2), 273–296. https://doi.org/10.1016/j.jecp.2013.02.004

Dunn, D. M. (2019). *Peabody Picture Vocabulary Test* (5th ed.). [Measurement instrument]. Minneapolis, MN: NCS Pearson.

Durso, F. T., & Shore, W. J. (1991). Partial knowledge of word meanings. *Journal of Experimental Psychology: General*, *120*(2), 190–202. https://doi.org/10.1037/0096-3445.120.2.190

Efron, B., & Tibshirani, R. J. (1994). *An introduction to the bootstrap*. Boca Raton, FL: Chapman and Hall/CRC. https://doi.org/10.1201/9780429246593

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Fernald, A., Zangl, R., Portillo, A. L., & Marchman, V. A. (2008). Looking while listening: Using eye movements to monitor spoken language comprehension by infants and young children. In I. A. Sekerina, E. Fernandez, &

H. Clahsen (Eds.), *Developmental psycholinguistics: On-line methods in children's language processing* (pp. 97–135). Amsterdam, The Netherlands: John Benjamins.

Fitneva, S. A., & Christiansen, M. H. (2011). Looking in the wrong direction correlates with more accurate word learning. *Cognitive Science*, *35*(2), 367–380. https://doi.org/10.1111/j.1551-6709.2010.01156.x

Fitneva, S. A., & Christiansen, M. H. (2017). Developmental changes in cross-situational word learning: The inverse effect of initial accuracy. *Cognitive Science*, *41*, 141–161. https://doi.org/10.1111/cogs.12322

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

Frederick Warne (188?). (n.d.). *The children's object book*. New York: F. Warne & Co. https://www.loc.gov/item/42048867/

Gershkoff-Stowe, L., & Hahn, E. R. (2013). Word comprehension and production asymmetries in children and adults. *Journal of Experimental Child Psychology*, *114*(4), 489–509. https://doi.org/10.1016/j.jecp.2012.11.005

Gillette, J., Gleitman, H., Gleitman, L., & Lederer, A. (1999). Human simulations of vocabulary learning. *Cognition*, *73*(2), 135–176. https://doi.org/10.1016/S0010-0277(99)00036-0

Golinkoff, R. M. & Hirsh-Pasek, K.(2000). *Becoming a word learner: A debate on lexical acquisition*. Oxford, England: Oxford University Press. https://doi.org/10.1093/acprof:oSo/9780195130324.001.0001

Gleitman, L. R., & Trueswell, J. C. (2020). Easy words: Reference resolution in a malevolent referent world. *Topics in Cognitive Science*, *12*, 22–47. https://doi.org/10.1111/tops.12352

Gurteen, P. M., Horne, P. J., & Erjavec, M. (2011). Rapid word learning in 13-and 17-month-olds in a naturalistic two-word procedure: Looking versus reaching measures. *Journal of Experimental Child Psychology*, *109*(2), 201–217. https://doi.org/10.1016/j.jecp.2010.12.001

Harris, J. (1862). *Marmaduke multiply*. New York: Blakeman & Mason. https://www.loc.gov/item/73169324/

Hendrickson, K., Mitsven, S., Poulin-Dubois, D., Zesiger, P., & Friend, M. (2015). Looking and touching: What extant approaches reveal about the structure of early word knowledge. *Developmental Science*, *18*(5), 723–735. https://doi.org/10.1111/desc.12250

Hoemann, K., Xu, F., & Barrett, L. F. (2019). Emotion words, emotion concepts, and emotional development in children: A constructionist hypothesis. *Developmental Psychology*, *55*(9), 1830–1849. https://doi.org/10.1037/dev0000686

Hopkins, C. B. (1919). *A dog's life by Tige: Army cartoons and comic history of our boys in France*. Toledo, OH: The Eoff-Sewell Company. https://www.loc.gov/item/20001390/

Horst, J. S., & Samuelson, L. K. (2008). Fast mapping but poor retention by 24-month-old infants. *Infancy*, *13*(2), 128–157. https://doi.org/10.1080/15250000701795598

Horst, J. S., & Houston-Price, C. (2015). Editorial: An open book: What and how young children learn from picture and story books. *Frontiers in Psychology*, *6*, 1719. https://doi.org/10.3389/fpsyg.2015.01719

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*(4), 434–446. https://doi.org/10.1016/j.jml.2007.11.007

Kachergis, G., & Yu, C. (2018). Observing and modeling developing knowledge and uncertainty during cross-situational word learning. *IEEE Transactions on Cognitive and Developmental Systems*, *10*(2), 227–236. https://doi.org/10.1109/TCDS.2017.2735540

Knabe, M. L., & Vlach, H. A. (2023). Not all is forgotten: Children's associative matrices for features of a word learning episode. *Developmental Science*, *26*(2), e13291. https://doi.org/10.1111/desc.13291

Kominsky, J. F., & Keil, F. C. (2014). Overestimation of knowledge about word meanings: The "Misplaced Meaning" effect. *Cognitive Science*, *38*(8), 1604–1633. https://doi.org/10.1111/cogs.12122

Kucker, S. C., McMurray, B., & Samuelson, L. K. (2015). Slowing down fast mapping: Redefining the dynamics of word learning. *Child Development Perspectives*, *9*(2), 74–78. https://doi.org/10.1111/cdep.12110

LaTourrette, A. S., & Waxman, S. R. (2020). Naming guides how 12-month-old infants encode and remember objects. *PNAS Proceedings of the National Academy of Sciences of the United States of America*, *117*(35), 21230–21234. https://doi.org/10.1073/pnas.2006608117

LaTourrette, A. S., Yang, C., & Trueswell, J. (2022). When close isn't enough: Semantic similarity does not facilitate cross-situational word-learning. In J. Culbertson, A. Perfors, H. Rabagliati, & V. Ramenzoni (Eds.), *Proceedings of the 44th annual meeting of the cognitive science society* (pp. 2554–2560). Cognitive Science Society: Toronto, ON.

Lo, C. H., Rosslund, A., Chai, J. H., Mayor, J., & Kartushina, N. (2021). Tablet assessment of word comprehension reveals coarse word representations in 18-20-month-old toddlers. *Infancy*, *26*(4), 596–616. https://doi.org/10.1111/infa.12401

MacNamara, J. (1972). Cognitive basis of language learning in infants. *Psychological Review*, *79*(1), 1–13. https://doi.org/10.1037/h0031901

Masek, L. R., Ramirez, A. G., McMillan, B. T. M., Hirsh-Pasek, K., & Golinkoff, R. M. (2021). Beyond counting words: A paradigm shift for the study of language acquisition. *Child Development Perspectives*, *15*(4), 274–280. https://doi.org/10.1111/cdep.12425

Mcloughlin Bro's. (1888). *The Circus procession*. New York: McLoughlin Bro's. https://www.loc.gov/item/2003612208/

McMurray, B., Horst, J. S., & Samuelson, L. K. (2012). Word learning emerges from the interaction of online referent selection and slow associative learning. *Psychological Review*, *119*(4), 831–877. https://doi.org/10.1037/a0029872

Medina, T. N., Snedeker, J., Trueswell, J. C., & Gleitman, L. R. (2011). How words can and cannot be learned by observation. *Proceedings of the National Academy of Sciences*, *108*(22), 9014–9019. https://doi.org/10.1073/pnas.1105040108

Montag, J. L., Jones, M. N., & Smith, L. B. (2015). The words children hear: Picture books and the statistics for language learning. *Psychological Science*, *26*(9), 1489–1496. https://doi.org/10.1177/0956797615594361

Nation, K., Dawson, N. J., & Hsiao, Y. (2022). Book language and its implications for children's language, literacy, and development. *Current Directions in Psychological Science*, *31*(4), 375–380. https://doi.org/10.1177/09637214221103264

Nayar, N., & Verulkar, N. (2018). *Hungry on the steps*. Pratham Books. https://freekidsbooks.org/hungry-on-the-steps-early-learning/

O'Rear, C. D., McNeil, N. M., & Kirkland, P. K. (2020). Partial knowledge in the development of number word understanding. *Developmental Science*, *23*, e12944. https://doi.org/10.1111/desc.12944

Oliva, A. (2005). Gist of the scene. In L. Itti, G. Rees, & J. K. Tsotsos (Eds.), *Neurobiology of attention* (pp. 251–256). New York: Academic Press.

Piccin, T. B., & Waxman, S. R. (2007). Why nouns trump verbs in word learning: New evidence from children and adults in the Human Simulation Paradigm. *Language Learning and Development*, *3*(4), 295–323. https://doi.org/10.1080/15475440701377535

Roembke, T. C., & McMurray, B. (2016). Observational word learning: Beyond propose-but-verify and associative bean counting. *Journal of memory and language*, *87*, 105–127. https://doi.org/10.1016/j.jml.2015.09.005

Roembke, T. C., Simonetti, M. E., Koch, I., & Philipp, A. M. (2023). What have we learned from 15 years of research on cross-situational word learning? A focused review. *Frontiers in Psychology*, *14*, 1175272. https://doi.org/10.3389/fpsyg.2023.1175272

Rowe, M. L., & Snow, C. E. (2020). Analyzing input quality along three dimensions: Interactive, linguistic, and conceptual. *Journal of Child Language*, *47*(1), 5–21. https://doi.org/10.1017/S0305000919000655

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

Saji, N., Imai, M., Saalbach, H., Zhang, Y., Shu, H., & Okada, H. (2011). Word learning does not end at fast-mapping: Evolution of verb meanings through reorganization of an entire semantic domain. *Cognition*, *118*(1), 48–64. https://doi.org/10.1016/j.cognition.2010.09.007

Shatz, M., Tare, M., Nguyen, S. P., & Young, T. (2010). Acquiring non-object terms: The case for time words. *Journal of Cognition and Development*, *11*(1), 16–36. https://doi.org/10.1080/15248370903453568

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

Smith, L. B., Suanda, S. H., & Yu, C. (2014). The unrealized promise of infant statistical word—Referent learning. *Trends in Cognitive Sciences*, *18*(5), 251–258. https://doi.org/10.1016/j.tics.2014.02.007

Smith, L. B., Jayaraman, S., Clerkin, E., & Yu, C. (2018). The developing infant creates a curriculum for statistical learning. *Trends in Cognitive Sciences*, *22*(4), 325–336. https://doi.org/10.1016/j.tics.2018.02.004

Smith, L., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, *106*(3), 1558–1568. https://doi.org/10.1016/j.cognition.2007.06.010

Smith, L. B., Yu, C., & Pereira, A. F. (2011). Not your mother's view: The dynamics of toddler visual experience. *Developmental Science*, *14*(1), 9–17. https://doi.org/10.1111/j.1467-7687.2009.00947.x

Snedeker, J., & Gleitman, L. R. (2004). Why it is hard to label our concepts. In D. G. Hall & S. R. Waxman (Eds.), *Weaving a lexicon* (pp. 257–293). Cambridge, MA: MIT Press.

Stevens, J. S., Gleitman, L. R., Trueswell, J. C., & Yang, C. (2017). The pursuit of word meanings. *Cognitive Science*, *41*, 638–676. https://doi.org/10.1111/cogs.12416

Suanda, S. H., Mugwanya, N., & Namy, L. L. (2014). Cross-situational statistical word learning in young children. *Journal of Experimental Child Psychology*, *126*, 395–411. https://doi.org/10.1016/j.jecp.2014.06.003

Swingley, D. (2010). Fast mapping and slow mapping in children's word learning. *Language Learning and Development*, *6*(3), 179–183. https://doi.org/10.1080/15475441.2010.484412

Tamis-LeMonda, C. S., Custode, S., Kuchirko, Y., Escobar, K., & Lo, T. (2019). 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

Tillman, K. A., & Barner, D. (2015). Learning the language of time: Children's acquisition of duration words. *Cognitive Psychology*, *78*, 57–77. https://doi.org/10.1016/j.cogpsych.2015.03.001

Trueswell, J. C., Lin, Y., Armstrong, B., III, Cartmill, E. A., Goldin-Meadow, S., & Gleitman, L. R. (2016). Perceiving referential intent: Dynamics of reference in natural parent–child interactions. *Cognition*, *148*, 117–135. https://doi.org/10.1016/j.cognition.2015.11.002

Trueswell, J. C., Medina, T. N., Hafri, A., & Gleitman, L. R. (2013). Propose but verify: Fast mapping meets cross-situational word learning. *Cognitive Psychology*, *66*(1), 126–156. https://doi.org/10.1016/j.cogpsych.2012.10.001

Twomey, K. E., & Westermann, G. (2018). Learned labels shape pre-speech infants' object representations. *Infancy*, *23*(1), 61–73. https://doi.org/10.1111/infa.12201

Wagner, K., Chu, J., & Barner, D. (2019). Do children's number words begin noisy? *Developmental Science*, *22*(1), 1–16. https://doi.org/10.1111/desc.12752

Wagner, K., Dobkins, K., & Barner, D. (2013). Slow mapping: Color word learning as a gradual inductive process. *Cognition*, *127*(3), 307–317. https://doi.org/10.1016/j.cognition.2013.01.010

Widen, S. C., & Russell, J. A. (2008). Children acquire emotion categories gradually. *Cognitive Development*, *23*(2), 291–312. https://doi.org/10.1016/j.cogdev.2008.01.002

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

Woodward, A. L., & Markman, E. M. (1998). Early word learning. In W. Damon (Ed.), *Handbook of child psychology: Vol. 2. Cognition, perception, and language* (pp. 371–420). Chichester: John Wiley & Sons, Inc.

Yoshida, H., & Smith, L. B. (2008). What's in view for toddlers? Using a head camera to study visual experience. *Infancy*, *13*(3), 229–248. https://doi.org/10.1080/15250000802004437

Yu, C., & Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science*, *18*(5), 414–420. https://doi.org/10.1111/j.1467-9280.2007.01915.x

Yu, C., & Smith, L. B. (2012a). Modeling cross-situational word–referent learning: Prior questions. *Psychological Review*, *119*(1), 21–39. https://doi.org/10.1037/a0026182

Yu, C., & Smith, L. B. (2012b). Embodied attention and word learning by toddlers. *Cognition*, *125*(2), 244–262. https://doi.org/10.1016/j.cognition.2012.06.016

Yu, C., Zhang, Y., Slone, L. K., & Smith, L. B. (2021). The infant's view redefines the problem of referential uncertainty in early word learning. *Proceedings of the National Academy of Sciences*, *118*(52), e2107019118. https://doi.org/10.1073/pnas.2107019118

Yurovsky, D., & Frank, M. C. (2015). An integrative account of constraints on cross-situational learning. *Cognition*, *145*, 53–62. https://doi.org/10.1016/j.cognition.2015.07.013

Yurovsky, D., Fricker, D. C., Yu, C., & Smith, L. B. (2014). The role of partial knowledge in statistical word learning. *Psychonomic Bulletin & Review*, *21*(1), 1–22. https://doi.org/10.3758/s13423-013-0443-y

Yurovsky, D., Smith, L. B., & Yu, C. (2013). Statistical word learning at scale: The baby's view is better. *Developmental Science*, *16*(6), 959–966. https://doi.org/10.1111/desc.12036

Zettersten, M., Wojcik, E., Benitez, V. L., & Saffran, J. (2018). The company objects keep: Linking referents together during cross-situational word learning. *Journal of Memory and Language*, *99*, 62–73. https://doi.org/10.1016/j.jml.2017.11.001

Zhang, Y., Chen, C. H., & Yu, C. (2019). Mechanisms of cross-situational learning: Behavioral and computational evidence. *Advances in Child Development and Behavior*, *56*, 37–63. https://doi.org/10.1016/bs.acdb.2019.01.001

Zhang, Y., Yurovsky, D., & Yu, C. (2021). Cross-situational learning from ambiguous egocentric input is a continuous process: Evidence using the human simulation paradigm. *Cognitive Science*, *45*(7), e13010. https://doi.org/10.1111/cogs.13010

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Supporting Information

Appendix

This appendix presents data for each of the eight individual target nouns. Included in the tables below are: (1) the mean proportion correct in WI test trials; (2) the total number of 2AFC trials that preceded an incorrect WI test trial (note: these are the trials that serve as the data for the current study's key partial learning analyses) and the breakdown of such trials that were first, second, or third in order; (3) the means and 95% confidence intervals (via bias corrected and accelerated bootstrap resampling with 5,000 samples) around means of participants' performance on the 2AFC test trials that preceded an incorrect WI test trial; the annotations reflect statistical significance levels of single-sample *t*-tests comparing mean 2AFC performance prior to an incorrect WI test trial against chance levels (0.50). Table A1 displays the by-target noun data for Experiment 1; Table A2 displays the same data for Experiment 2.

Table A1
By-target noun descriptive statistics and partial learning in Experiment 1

Target Noun	Word Identity Test (*SD*)	Partial Learning				
		Total Trials (Participants [a])	First Trials	Second Trials	Third Trials	2AFC (Bias-Corrected and Accelerated [BCa CI])
Apple	0.57 (0.49)	77 (27)	25	26	26	0.83 [0.73–0.90]***
Bird	0.59 (0.46)	73 (29)	23	27	23	0.63 [0.49–0.73]*
Book	0.86 (0.33)	26 (11)	9	8	9	0.71 [0.47–0.86]^
Dog	0.44 (0.49)	100 (35)	32	35	33	0.66 [0.58–0.74]***
Door	0.70 (0.45)	54 (19)	17	19	18	0.65 [0.49–0.77]*
Flower	0.54 (0.47)	81 (31)	25	27	29	0.70 [0.58–0.80]**
Hat	0.51 (0.50)	89 (31)	30	30	29	0.53 [0.42–0.62][ns]
Shoe	0.31 (0.44)	125 (44)	39	44	42	0.67 [0.59–0.74]***

Note. [a]The number of unique participants that contributed 2AFC trials that preceded an incorrect WI test trial. Because each participant saw three AFC trials per target noun, the number of first, second, and third trials per target noun is equivalent to the number of unique participants who contributed. Thus, only one number is presented for first, second, and third trials.

*$p < .05$; **$p < .01$; ***$p < .001$; ^$p < .10$; $^{ns}p > .10$.

Table A2
By-target noun descriptive statistics and partial learning in Experiment 2

Target Noun	Word Identity Test (*SD*)	Partial Learning				
		Total Trials (Participants)	First Trials	Second Trials	Third Trials	2AFC [BCa CI]
Apple	0.34 (0.46)	118 (42)	40	39	39	0.84 [0.75–0.91]***
Bird	0.19 (0.39)	146 (50)	49	48	49	0.77 [0.70–0.83]***
Book	0.38 (0.48)	111 (38)	38	37	36	0.53 [0.41–0.63][ns]
Dog	0.23 (0.43)	138 (46)	46	46	46	0.54 [0.46–0.62][ns]
Door	0.28 (0.42)	130 (46)	43	43	44	0.68 [0.59–0.75]***
Flower	0.46 (0.48)	98 (35)	31	33	34	0.64 [0.52–0.74]*
Hat	0.39 (0.48)	109 (38)	36	36	37	0.63 [0.52–0.74]*
Shoe	0.08 (0.28)	165 (55)	55	55	55	0.65 [0.59–0.72]***

Note. *$p < .05$; **$p < .01$; ***$p < .001$; ^$p < .10$; $^{ns}p > .10$.