*Developmental Science*

WILEY

Developmental Science



RESEARCH ARTICLE OPEN ACCESS

# The Unforgettable "Mel": Pragmatic Inferences Affect How Children Acquire and Remember Word Meanings

Katherine Trice[1] | Dionysia Saratsli[2] | Anna Papafragou[3] | Zhenghan Qi[1,4]

[1]Department of Psychology, Northeastern University, Boston, Massachusetts, USA | [2]Department of Linguistics and Cognitive Science, University of Delaware, Newark, Delaware, USA | [3]Department of Linguistics, University of Pennsylvania, Philadelphia, Pennsylvania, USA | [4]Department of Communication Sciences and Disorders, Northeastern University, Boston, Massachusetts, USA

**Correspondence:** Katherine Trice (trice.k@northeastern.edu)

**Received:** 14 August 2024 | **Revised:** 9 February 2025 | **Accepted:** 5 March 2025

**Funding:** This research was supported by internal funds from the University of Delaware awarded to Zhengan Qi.

**Keywords:** informativeness | pragmatics | scalar implicature | theory of mind | word learning

## ABSTRACT

Children can acquire novel word meanings by using pragmatic cues. However, previous literature has frequently focused on in-the-moment word-to-meaning mappings, not delayed retention of novel vocabulary. Here, we examine how children use pragmatics as they learn and retain novel words. Thirty-three younger children (mean age: 5.0, range: 4.0–6.0, 21 girls; 85% White) and 33 older children (mean age: 7.5, range: 6.1–9.2, 16 girls, 66% White) participated. During learning, the sound-meaning mapping was either readily available (Direct Mapping condition) or required pragmatic inference (Inference condition). Children's word retention was tested immediately after learning and after 10–15 min of delay. Across both conditions, children performed similarly during learning. There were no significant differences between conditions for either immediate recall or retention in younger children. Importantly, retention (but not immediate recall) in older children demonstrated a significant advantage for the Inference over the Direct Mapping condition. Word retention in the Inference condition was predicted by age and mediated by children's ToM ability. We conclude that children can successfully acquire and retain meanings via pragmatic inference; moreover, the effects of active pragmatic computation on meaning retention grow with development. Such a developmental difference in meaning consolidation is possibly mediated by children's developing ToM skills.

## 1 | Introduction

### 1.1 | Word Learning and Pragmatic Inference

Word learning is fundamentally a task of discovering meanings. But sound-meaning mappings are not always readily available in the context, and young learners must use a variety of linguistic and extralinguistic cues to narrow the referential space for novel words in their input (Fisher et al. 2010; Gleitman 1990; Halberda 2003, 2006; Naigles 1990). In so doing, learners often draw pragmatic inferences as to the intended meaning of a new word (Gollek and Doherty 2016; Frank and Goodman 2014; for a review, see Papafragou and Grigoroglou 2021).

Pragmatic inferences are based on the communicative principles used by a speaker within a conversation. One such principle relates to informativeness and posits that cooperative speakers should strive to offer as much information as is relevant to a dialogue (Grice 1975). Consider a scene with two toy dinosaurs: both dinosaurs have a shared feature—for example, a bandanna—and one has an additional, unique feature—for example, a headband (Frank and Goodman 2014). An experimenter points to the toy with the two features and says, "This is a dinosaur with a dax!" Children as young as 3 understand that "dax" must refer to the unique feature (the headband), presumably because they assume that the speaker intended to supply enough information to secure reference and choosing the shared feature (the bandanna) would fail to fulfill the requirement to be informative. Notably, such referential ambiguity cannot be resolved solely based on the linguistic—lexical, syntactic, or semantic—information regardless of visual context, as both bandanna and headband can be referential candidates for *dax*. Instead, abstract reasoning about the speaker's intentions (including the informativeness and relevance of what was said) is a necessary level of linguistic interpretation. Despite this and other evidence of early sensitivity to informativeness (Stiller et al. 2015; Kampa and Papafragou 2020, 2023; Nadig and Sedivy 2002), children do not always apply this principle to resolve meaning ambiguities in an adult-like way before 6 years of age (Robinson and Whittaker 1985; Feeney et al. 2004; Katsos and Bishop 2011; Barner et al. 2011; Grigoroglou and Papafragou 2019a, 2019b, among others).

**Summary**

- Children as young as 4 can reliably apply assumptions about the choices and intent of others to resolve ambiguities in speech.
- This process, supported by social cognition, may lead to lasting memory of novel words in children 6 and up.
- These findings reveal an increasingly important role of social cognition in building children's vocabulary as they enter school-age years.

## 1.2 | Pragmatic Inference and Memory for Word Meanings

There has been extensive literature studying word learning in children across a wide variety of scenarios. Much of this has focused on in-the-moment meaning mappings or recall tested directly after a meaning-mapping phase (Bion et al. 2013; Markman and Wachtel 1988; Vlach and DeBrock 2017; Zosh et al. 2013). However, because of methodological constraints, comparably less work has used measures of delayed meaning retention. This can pose a critical gap in the literature, as a successful initial mapping does not always transition into long-term retention (Axelsson et al. 2012; Horst and Samuelson 2008; Bredemann and Vlach 2021). Instead, long-term retention necessitates the engagement of memory-supporting processes during and/or after learning, with more numerous and powerful memory supports within a given learning scenario resulting in better long-term retention (Vlach and Sandhofer 2012). Substantive literature has sought to fill this gap by exploring word learning in specific scenarios and even relating it to possible underlying mechanisms such as visual and auditory attention, visual recognition memory, visual cue use, and saliency (Kucker and Samuelson 2012; Pino Escobar et al. 2023; Hartley et al. 2020; Vlach and DeBrock 2019). But little is known about how the development of social cognitive skills interact with vocabulary acquisition, specifically in its retention phase, and the role that pragmatic computation may play within this.

Of the comparatively more limited studies that have investigated word retention, we focus on those that have indicated that pragmatic inference may play an important role. For example, Zosh et al. (2013) studied 3-year-olds' preference to map an ambiguous novel word onto a novel object in a scene instead of a known object with a familiar name, a pattern known as *mutual exclusivity*. Importantly, children were more successful in immediately retrieving words inferred through mutual exclusivity than those acquired through direct instruction (or direct mapping). These results open up clear avenues for further research. First, since immediate retention occurred—as in many such experiments—only a few minutes at most after the initial mapping, it is currently unclear how stable the memories of these words are. Work such as Carey and Bartlett (1978) do indicate that words mapped in a situation involving mutual exclusivity can be remembered longer term; however, it lacks a contest to determine the relative effect that word learning has via this route. This is further supported by work such as Pino Escobar et al. (2023), which found better learning, immediate retention, and delayed retention for words mapped via mutual exclusivity and from eBook reading (which included characters often engaged in social situations) than those learned statistically via cross-situational word learning. However, none of these works sought to explore the role of pragmatic inference directly or how social cognitive skills and use may interact with this.

Studies such as Zosh et al. (2013) raise (but do not resolve) two possibilities for the role of inferential computations in word retention. A first possibility is that inferences facilitate word retention via active learning and encoding. We know that words learned in scenarios where children can manipulate order and pacing have an advantage in long-term memory retention compared to passively learned words (Ruggeri et al. 2019). In Zosh et al. (2013), the need for more active cognitive control to compare the novel word to a known competitor in the inference condition may have engaged active learning. However, evidence from the study does not uniquely lead to this interpretation because children were also asked to point to the target object in the inference condition, potentially boosting the robustness of the new word, while no action was called for in the more passive instruction condition.

A second, *non-disjoint* possibility is that making inferences through mutual exclusivity might involve heavier engagement of *Theory of Mind* (ToM), the ability to reason about others' minds (Baron-Cohen et al. 1985, 1997; Wellman et al. 2001). ToM might be associated with richer and deeper semantic processing during encoding (Li and Jeong 2020). If reasoning about the speaker's mental state and knowledge is part and parcel of mutual exclusivity (see, e.g., Diesendruck and Markson 2001; cf. Grice 1975; Sperber and Wilson 1986), then ToM would support pragmatic inferences leading to the correct word-referent assignment and, thereafter, to long-term knowledge of the newly-heard word. This may be further supported by Pino Escobar et al. (2023)'s findings, as the advantage of mutual exclusivity and eBook reading may stem in part from the pragmatic and social aspects of these—in the latter case, the eBook often included characters engaged in social situations and interactions when the novel words were raised. This possibility is consistent with work beyond the narrow case of mutual exclusivity (as a type of pragmatic inference) pointing to close links between various pragmatic phenomena and ToM abilities in both adults and children (Fairchild and Papafragou 2021; Gollek and Doherty 2016; Matthews et al. 2018; Brosseau-Liard et al. 2015; Sabbagh and Baldwin 2001). For instance, Fairchild and Papafragou (2021) showed that individual adults' ToM differences uniquely predicted their pragmatic reasoning, with stronger ToM skills predicting better assessments of scalar implicatures and indirect requests. Furthermore, this possibility is supported by neuroimaging evidence of ToM involvement in pragmatic inference, as brain regions associated with ToM, such as right Temporo-Parietal Junction (rTPJ) and medial prefrontal cortex (mPFC) (e.g., Saxe and Kanwisher 2003), are activated during the processing of pragmatic phenomena (e.g., Bambini et al. 2011; Bašnáková et al. 2014; Hornick and Shetreet 2022). However, a possible role of ToM in the retention advantage of newly-acquired words is challenged by proposals contesting the involvement of ToM both in the specific case of mutual exclusivity (Srinivasan et al. 2019; de Marchena et al. 2011; Preissler and Carey 2005) and in other pragmatic phenomena (Geurts et al. 2019; Andrés-Roqueta and Katsos 2017). In these proposals, the involvement of ToM is neither consistent nor necessary during pragmatic computations. For instance, what might appear to be effects of speaker meaning when taking into account informativeness might instead arise from a comparison of what the speaker has said to other possible lexical alternatives without considering the speaker's intention or mental state in any significant sense. In support of this possibility, autistic individuals succeed with mutual exclusivity despite having difficulties computing others' beliefs in independent ToM tasks (Preissler and Carey 2005; de Marchena et al. 2011). Furthermore, as seen in work such as Hartley et al. (2020), a mere presence of a low-level social cue that may involve ToM reasoning (such as a head-turn or eye gaze for referential disambiguation) does not always result in better retention than a non-social cue. However, it should be noted that such low-level cues develop and quickly become automatic early in infancy, and thus, social facilitatory effects may be limited as one develops (Çetinçelik et al., 2021). At present, then, the role of ToM for inferential computations during word learning requires more experimental attention.

### 1.3 | Current Study

Here, we aim to address the role of pragmatics and ToM in how children learn novel word meanings. We take "learning" to extend beyond immediate mappings and view it through the lens of a multi-stage encoding and consolidation process that extends to longer-term retention. We contrast two learning contexts: a context that provides a direct mapping between a novel label and the target novel object, and a context in which abstract pragmatic principles such as informativeness are required in order to infer the label-object link. In a prior version of this paradigm with adults (Trice et al. 2023), novel words learned through pragmatic inference were better retained than those learned through direct mappings; furthermore, retention in the inferential cases was associated with individuals' ToM scores. In the current paper, we seek to determine (a) whether children can use pragmatics to infer word meanings, and if so, whether there is a retention advantage for pragmatically inferred words and (b) if and how this advantage is intertwined with the development of ToM skills. We compare a younger (4–6-year-old) and an older (7–9-year-old) child group. This age split was implemented because, as mentioned, informativeness-based inferences may not become reliable across the board until the age of six (Robinson and Whittaker 1985; Feeney et al. 2004; Katsos and Bishop 2011; Barner et al. 2011; Grigoroglou and Papafragou 2019a, 2019b, among others). Moreover, ToM skills are developing in sophistication across this age range on both a behavioral and neurological level (Gweon et al. 2012; Richardson et al. 2018, Wellman and Liu 2004).

We expect that during mapping via pragmatic inference, learners must construct an internal model of the speaker and the other alternatives they could have used, determine their intent based on the limited input given, and use that intent to select the correct referent—all of which require greater internal computation and judgement than direct mapping—and hence greater control over and engagement with their cognitive processes. Thus, we hypothesize that active encoding would emerge under these conditions, leading to stronger memory encoding of the word-object representations computed via pragmatic inference. As these mechanisms require ToM, we predict that maturing ToM skills might underlie an increase in memory advantage for inferred over directly mapped words independent of a general increase in vocabulary retention across conditions with age. Finally, we hypothesize that younger children will show less of an advantage for word learning that requires pragmatic inference as opposed to direct mapping compared to older children. Note that if both a pragmatic inference advantage and a relation to ToM are seen, this experiment cannot fully differentiate a mechanistic account where stronger memory encoding is related to emerging advantages of active encoding via mentalizing from another account containing a specific interrelation of ToM use and memory. As such, both will be discussed as relevant.

## 2 | Experiment

### 2.1 | Participants

Families were recruited via study posts at ChildrenHelpingScience.com and the Language Research Registry at the University of Delaware. All children (4.0–9.2) were native speakers of English with no extensive exposure (defined as more than 10 h of exposure per week) to other languages before the age of 5. All parents/guardians gave informed consent, and children assent, prior to inclusion. Children were split into two groups: younger children (4.0–6.0 years old) and older children (6.1–9.2 years old, Table 1). As discussed previously, this split is based on prior research suggesting that informativeness-based inferences may not be reliable until the age of 6 (see Sections 1.1 and 1.3). A stopping rule of 1 year of data collection with equal-sized groups of at least 30 was implemented to align well with prior adult work in this paradigm (Trice et al. 2023). Based on prior work, this would ensure a power of, at minimum, 75% for our between-group comparisons and 80% for our across-group individual difference analysis. No participants were excluded. The two groups were similar in terms of gender and race.

### 2.2 | Stimuli and Procedure

The experiment was administered via Zoom. Children performed the tasks in their home environment. The procedure was divided

**TABLE 1** | Participant demographics by age group.

| Variable | Younger children ($n$ = 33, 21 girls, 85% White) | | | Older children ($n$ = 33, 16 girls, 66% White) | | | $p$ |
|---|---|---|---|---|---|---|---|
| | Mean | SD | Range | Mean | SD | Range | |
| Age (years) | 5.0 | 0.6 | 4.0–6.0 | 7.5 | 0.9 | 6.1–9.2 | <0.01 |
| ToM accuracy | 0.57 | 0.21 | 0.15–0.96 | 0.81 | 0.15 | 0.39–1.00 | <0.01 |
| Flanker score (ms) | 756.43 | 1386.31 | −558.70 to 4943.39 | 412.52 | 1272.80 | −663.10 to 6669.63 | 0.31 |
| Control question accuracy | 0.87 | 0.19 | 0.33–1.00 | 0.94 | 0.16 | 0.33–1.00 | 0.06 |

*Note:* The Flanker Score was calculated via reaction time differences between the incongruent and congruent trials. Control question accuracy is based on the score for the selected control questions in the ToM task. The gender ratio did not differ significantly between groups ($p$ = 1). The Flanker score showed a significant congruency effect within each group (1-tailed $t$-test: $p_{\text{Older}} = 0.04$, $p_{\text{Younger}} = 0.002$).
Abbreviations: ms, milliseconds; SD, standard deviation.

into the following ordered phases: a word learning task (6 min), a ToM task (15 min), a word retention task (1 min), and an executive function (EF) task (Flanker, 4 min). The word learning task, word retention task, and Flanker task were hosted on the Gorilla Experimental Builder (Anwyl-Irvine et al. 2020). The ToM task was administered using PowerPoint.

#### 2.2.1 | Word Learning Task

Our task featured two word learning conditions, Inference and Direct Mapping, in a within-subject design (Figure 1). Children were introduced to a cartoon girl, Mary, who would show them her favorite toys. Word learning began with two practice trials to ensure task comprehension. Here, children were briefly introduced to a dinosaur and an alien toy holding known objects (e.g., pencil and guitar in Figure 1A.1). One practice trial mirrored the setup of the Inference condition, where the mini-objects were used to disambiguate the identical figures, and the other mirrored the setup of the Direct Mapping condition, where the figures were used to disambiguate the referred mini-objects (Figure 1). Feedback was given.

Next, children began the learning trials, which featured four novel words per condition, with two trials per word. This number was chosen to ensure the task demand was reasonable for our youngest participants while still retaining the possibility of capturing within-participant variability. The novel words and novel objects were unique to each condition. Additionally, four filler trials with known object features were included in each condition as attention checks and to reinforce the practice trials; these mirrored the practice trial for each condition. The two conditions were manipulated in a counterbalanced blocked design—one block per condition—for a total of eight novel words (Figure 1). No feedback was given.

In the Inference condition, successful sound-meaning mappings required pragmatic inference. The visual stimuli contained two identical cartoon figures both holding a shared object (e.g., a rainbow toy in Figure 1A.2), with one of the figures also holding a unique object (e.g., a pink c-shaped object in Figure 1A.2). Here, Mary would say, for instance, "Look! I like this dinosaur; it is holding a *mel*!". The participant would then be asked which toy Mary liked. Children could not semantically resolve the ambiguity, as the novel word could refer to either target or competitor. To identify the correct dinosaur, the listener had to assume that the speaker had offered an appropriate amount of information in using the novel word to allow pragmatic disambiguation between the two dinosaurs. Thus, "mel" should be the unique novel object—the pink c-shaped object—and the participant should select the figure holding that object. Children's success in choosing the right referent was used as a measure of in-the-moment mapping for the Inference condition.

In the Direct Mapping condition, the novel word could be identified relatively straightforwardly from a sentence in combination with the scene, with little or no ambiguity. The visual stimuli for the Direct Mapping condition contained two different cartoon figures holding a shared novel object (e.g., an orange y-shaped object in Figure 1A.3). Here, Mary would say, for instance, "Look! I like this *bink*! It is on the alien!". In this case, the mapping between the novel word *bink* and the Y-shaped novel object was readily available for listeners.

Our design of the Inference condition was inspired by Frank and Goodman (2014) but differed from that study in two ways. First, we used completely novel objects to avoid any possible confounds from existing knowledge of object names, which might prevent children from assigning new labels. Additionally, having abstract novel objects for both the target and competitor allows us to circumvent the need to examine whether the presence of known competitors would lead to faster integration of a novel referent into the memory system (as found in Coutanche and Thompson-Schill 2014), thus contributing to our retention rates. Second, our design removed the pointing feature indicating the figure being discussed, a strong pragmatic visual cue helping narrow down the intended referent of the novel label.

At the end of each learning block, children's word retention was tested in a four-alternative-forced-choice task. Children saw four novel objects on the display and heard the question "Which one is a [novel-word]?". In each case, all three competitors came from the same condition as the target object (Figure 1A.4).

Children's word retention was tested again approximately 15 min after the end of the learning trials, after words have begun to move into long-term retention (Radvansky et al. 2022), in the same four-alternative-forced choice task. The order of the eight novel words

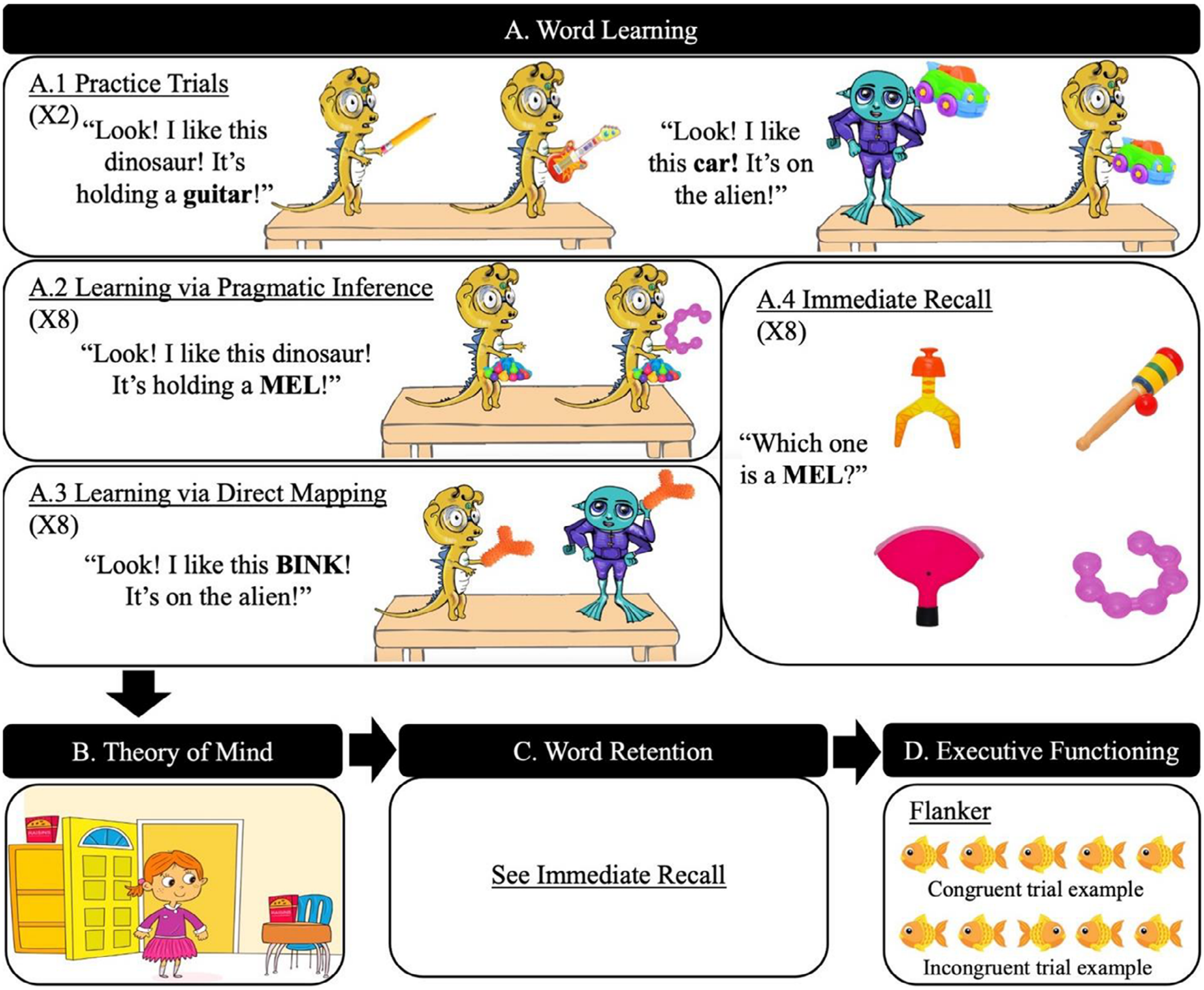


**FIGURE 1** | Experimental procedure and stimulus examples. The figure demonstrates example practice, word learning (direct mapping/inferential), and immediate recall trials, as well as an example of the ToM and Executive Functioning (Flanker) task visual stimuli. Task order was as shown. Text in quotations would be spoken aloud.

was randomized, and competitors for each trial were randomly selected across the two conditions with at least one competitor from the given target's condition (Figure 1).

To increase image saliency within the task's visual setting, all the pictures were edited to achieve brighter colors and higher saturation. Auditory stimuli were recorded in a quiet room by two female native speakers of English, one acting as the narrator and the other as Mary. Sound files were edited to remove any background noise, and identical intensity levels were set at 70 dB across all sentences. All sound edits were performed in *Praat* (Boersma and Weenink 2018).

### 2.2.2 | Individual Differences Tasks

**2.2.2.1 | ToM Task.** Between immediate recall and retention tests, children completed a ToM task modified from the picture booklet task in Richardson et al. (2018) (Original: https://osf.io/G5ZPV/). In this task, children listened to stories and answered questions after each story. For instance, as in Figure 1B, children were told that the illustrated girl had just entered the classroom and said, "I want that snack." They had to identify the box of raisins the girl wanted and how they knew she wanted those; this tested their use of social-cognitive cues such as eye-gaze and perspective-taking. As this task was highly verbal, control questions that required referential disambiguation but no ToM reasoning—relying solely on language comprehension and recall—were selected to examine children's baseline comprehension skills. Due to the time constraints of this online experiment, we shortened this task to eleven interactive stories and presented the stimuli on PowerPoint for remote administration. The 11 stories were specifically selected to ensure that they reflected skills developing across our age range—for instance, simpler false belief judgements should be emerging and improving across the age range of our younger children, while moral judgement based on false belief should still be in development for much of our older age range (Cushman et al. 2013; Wellman et al. 2001). We also included the stories that probed children's ability to understand others' preferences that differed from their own as a proxy for

ToM, as these skills are a precursor for false belief (Wellman and Liu 2004). This allowed us to have a more holistic view of ToM compared to other assessments. Finally, we ensured that each selected story could be modified for remote administration, as the original versions had children directly interacting with physical characters and scenery.

Children answered a total of 26 ToM questions and 4 generic comprehension questions, along with several unscored questions used to set up and transition within stories. The proportion of children's correct responses to the ToM questions was computed and used as our main ToM measure. Of the comprehension questions, three questions required participants to select between two referents and were thus selected as our control questions. Since only three questions were chosen, non-parametric tests were used for between-group comparisons.[1]

The children's behavior was double-scored by two trained research assistants based on the Zoom recordings. In cases where these scores differed on any given question, a third coder (the first author) would watch the video and determine the final score for that trial. Descriptive statistics of the scores for both the ToM questions and the control questions are available in Table 1.

**2.2.2.2 | Executive Functioning (Control) Task.** The experiment ended with an assessment of children's EF skills, specifically inhibitory control, via a Flanker Task. This task provided a control measure, as resolving referential ambiguity has been linked to EF (e.g., Qi et al. 2020; Woodard et al. 2016). In addition, ToM and EF are often correlated in the context of pragmatic processing, and EF can be involved in the ToM task itself (Apperly 2012; Carlson and Moses 2001; Fairchild and Papafragou 2021; Baillargeon et al. 2010; Scott et al. 2012; Setoh et al. 2016). We included this task to test whether any age effects in word learning could be uniquely explained by ToM.

The Flanker task used was adapted from the built-in materials available in the Gorilla task inventory (Rueda et al. 2004; Massonnié et al. 2019). This task used cartoon fish in place of the more traditional arrows to make it more engaging and child friendly. In this task, participants were presented with five fish, either looking all toward the same direction (left or right; congruent trials) or with the center one looking in the opposite direction (incongruent trials) and were instructed to press the button that matched the direction that the fish in the middle was facing, thereby inhibiting the direction of the surrounding fish. Children were first introduced to the task and familiarized with the procedure and buttons to press, with the instructions being read aloud. This was followed by four practice trials—two congruent, two incongruent—with feedback given to ensure that the child understood what key to press. Twenty-four trials were presented in randomized order, evenly divided between congruent and incongruent trials. Each trial would be preceded by a fixation cross that was presented in the middle of the screen for one at three different duration times: 400, 800, or 1200 ms. Each duration occurred eight times, four per trial type. We used the difference between participants' average reaction times for incongruent and congruent trials as a measure of their performance. Flanker reaction times less than 150 ms were dropped from analysis, as it takes more than 200 ms to plan an action after visual processing (Krigolson et al. 2015). These represented 2.4% of the data. Descriptive statistics of the Flanker scores can be seen in Table 1.

## 2.3 | Analysis Approach and Models

### 2.3.1 | Data Processing

Data were compiled for analyses into two different formats. The first was trial-by-trial binomial accuracy data for analyses on word learning outcomes (1 for correct and 0 for incorrect): initial mapping, immediate recall, and retention. The second was subject-level aggregated data for each learning outcome, ToM, and EF. For a handful of participants whose data were not correctly preserved for some sections on Gorilla, hand-scored from our recorded videos for accuracy on a trial-by-trial basis and saved as an Excel file. A separate Excel file was made for the ToM task containing the relevant scores. All subsequent data processing and organization were done within R (R Core Team 2021).

For our analyses on word learning, an additional column marking whether a given word was initially learned was appended. This was determined based on the initial mapping accuracy across its two learning trials. If both trials were answered correctly, the item would be marked with a 1 for learned. Otherwise, the item would be marked with a 0 for not learned. We chose this stricter cutoff to limit the potential impact of guessing rather than learning in our evaluation.

For our analyses on individual differences, accuracy was computed across trials for each session for each condition for each subject, along with each child's age, ToM task scores, and the RT difference scores between the congruent and incongruent Flanker trials. Gender was included as demographic information, but because it did not interact with other factors within our models, it was dropped from our final analysis outputs.

As video recordings were available in the majority of the participants who had compatible hardware devices, a decision was made post hoc to analyze the webcam eye tracking data (sampled at 30 Hz) during learning to investigate the online processing of pragmatic cues. In total, 21 older children and 24 younger children contributed data for this analysis. Data were collected, coded, and processed by the same trained individuals in the same way. Because of the post-hoc nature of this analysis, the comparison between older and younger children in the temporal course of pragmatic computation should be taken as simply suggestive evidence for developmental changes in online pragmatic processing.

### 2.3.2 | Data Analysis

**2.3.2.1 | Analysis of Word Learning Outcomes.** To examine children's performance during initial mapping, immediate recall, and retention, we constructed separate generalized binomial linear mixed models (*glmer*) for analysis of each learning outcome using the lme4 package of R (Baayen 2008; Baayen et al. 2008).

For initial mapping, older and younger participants were combined into a single data frame for analysis. In order to determine if mapping accuracy differed by age and whether an age effect was particularly salient in the Inference condition, we focused on the main effect of age group and the interaction between age group and condition on initial mapping accuracy. As our two conditions differed so strongly in what was asked (i.e., resolving a complex pragmatic inference in the Inference condition vs. simply selecting the novel object held by the named character in the Direct Mapping condition), the main effect of condition was deemed largely a byproduct of the model and, thus, theoretically unimportant. By-subject random intercepts and random slopes for condition and by-item random intercepts were used.

For immediate recall and retention, we assume fair comparison between the Inference and the Direct Mapping conditions as the four-alternative-forced-choice task was identical across conditions. Three models were constructed for each of the two learning outcomes. First, we tested the main effect of condition (Direct Mapping vs. Inference) with initial learning success (0 vs. 1) as a covariate on the trial-by-trial accuracy for each age group. By-subject random intercepts and random slopes for condition and by-item random intercepts were used. Then, we combined older and younger participants into a single data frame for analysis of the continuous age effect. We tested the main effects of age and condition, as well as the interaction between the two, along with the main effect of initial learning success on children's immediate recall or retention accuracy. In addition to the random slopes and intercepts of the separate group models, by-item random intercepts were used. Across all the generalized binomial linear mixed models, contrast coding of −0.5 versus 0.5 was established for each main effect in order to use the mean across the different levels for the given effect as the baseline. In each of the above analyses, the mean and confidence intervals for our groups by conditions were also calculated and compared to chance for the given timepoint to check if each performed above, at, or below chance. Results can be found in Section 2.5.1.1

A supplementary analysis examining change from immediate recall to retention was completed using the same setup of the three models for immediate learning and retention described above but with both testing phases included and phase as a main effect and interaction. The results can be found in Supporting Table A1.

**2.3.2.2 | Analysis of Individual Differences in Word Learning.** We also examined whether word learning, particularly when pragmatic inferences are involved, is uniquely related to ToM development. We first examined the normality of our dependent variables using quantile-quantile plots (Supporting Figure A1). If the data fell largely within the acceptable prediction interval surrounding the reference line, Pearson correlations would be used where appropriate; if not, Kendall correlations would be used. Additionally, because of the proportional nature of our dependent variables, which is especially clear for immediate recall, where each condition has only five possible accuracy values because only four novel words were learned per condition, both multiple linear regression models (*lm*) and generalized quasibinomial probit linear regression models (*glm*) were considered. Model selection was based on the root mean square errors of model fit.

A separate model was constructed for each learning outcome and for each condition. Children's performance in answering the ToM questions, control questions, Flanker task, and age served as the predictors for word learning. Additionally, for each learning outcome, the learning measure in prior sessions would be included as a covariate. Thus, a given immediate recall model would have learning accuracy as a predictor, and a given retention model would have both learning and immediate recall accuracy as predictors. Both general and generalized linear models for each case differed only in internal setup by the latter, including our quasibinomial family and probit link specifications. Root mean square errors (*rmse*) were used to compare each model to its competitor (Supporting Table A2).

For any significant predictors in the final models of selection, partial correlations (*pcor.test*) would then be run between the dependent variable of the model and the significant predictor while controlling the other key individual difference measures. If any of these came out to be no longer significant, a mediation analysis would then be run to see if the relationship was mediated by the controller (R Core Team 2021; Revelle 2023). The results of these analyses can be found in Section 2.5.2.

**2.3.2.3 | Analysis of Eye-Tracking Data.** Eye gaze was segmented during learning beginning 0.2 s before and ending 1.8 s after the onset of the final word in each condition and hand-coded frame-by-frame for fixation to target, to competitor, away, or un-codable eye-gaze. Any time points marked as un-codable for a given participant and trial were dropped from further analysis (9.5% of the data). Codable fixation data were divided into time bins with 8 observations per bin to allow the computation of the proportion of target/competitor fixations.

An analysis of the full time-window was used to determine whether looks-to-the-target and looks-to-the-competitor significantly differed for each group and, more critically, if this differed between groups. Online eye-tracking studies have shown comparable results in capturing group-level fixation patterns as their in-lab counterparts (Ovans et al. 2023; Kandel and Snedeker 2024). This was done using generalized binomial linear mixed models (*glmer*) in R, with one combined model and an additional model for each group. We tested the main effects of looks-to-object versus competitor and group, as well as the interaction between the two, for the combined model and just looks-to-object versus competitor for each individual model. By-subject random intercepts and random slopes for looking location and by-item random intercepts were used. Contrast coding of −0.5 versus 0.5 was once again used.

To capture the coarse timeline of referential disambiguation, a divergence-point analysis was used to determine where proportion of looks-to-the-target diverged from proportion of looks-to-the-competitor and whether the timing of divergence points significantly differed between groups for the inference condition. We applied a moving window width of 33 ms to capture the full timescale. As each participant had 8 items per condition, this allowed for approximately 64 observations per bin for a given participant. This analysis was done via the R eye-tracking analysis package stats. VWP (Ito and Knoeferle 2022). Here, 1000 iterations of bootstrapping were calculated. In each case, the log of the fixation proportion for the given time bin per participant

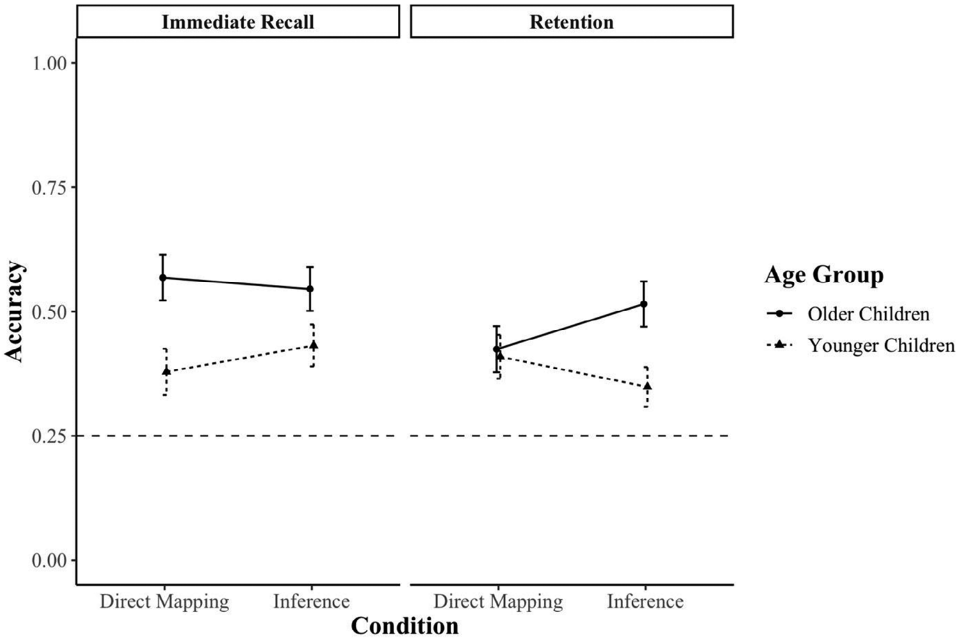


**FIGURE 2** | Accuracy-by-condition for older versus younger children for immediate recall and retention. Points represent mean accuracy. Error bars represent within-subject standard error. The horizontal dashed line at 0.25 represents chance.

per trial was calculated for the target and competitor, and a *t*-test was performed to see if this differed significantly at a given timepoint for a given group. The divergence point was determined to occur at the earliest point where ten consecutive time bins were significantly different. The ten consecutive time bins ensure independent observations between the first bin and the final bin. Confidence intervals of 95% for the mean divergence point for each group were calculated. To determine if the divergence points significantly differed between groups, the data were pooled and randomly assigned labels in another 1000 iterations of our bootstrapping to form a null distribution. The *p* value was calculated based on the proportion of occurrences in this null distribution that were larger than the differences in divergence points from our true data.

## 2.4 | Transparency and Openness

All data, preprocessing and analysis code, stimuli, and research materials are publicly available at https://osf.io/j5e9p/?view_only=b71f338bdb644c1da6c72137831a8778. Analysis packages can be found in Section 2.3. The study design and analyses were not pre-registered.

## 2.5 | Results

### 2.5.1 | Analysis of Word Learning by Group

**2.5.1.1 | Behavioral Accuracy Analysis.** First, we examined children's performance during the initial sound-meaning mapping. Both age groups performed significantly above chance across both the Direct Mapping and Inference conditions (chance: 0.50; Direct Mapping: younger children: 0.95 [$CI_{95\%}$ = 0.03]; older children: 0.98 [$CI_{95\%}$ = 0.03]; Inference: younger children: 0.70 [$CI_{95\%}$ = 0.05]; older children: 0.65 [$CI_{95\%}$ = 0.06]). Critically, there were no significant differences between the two age groups, nor an interaction between condition and age (group: estimate = 0.44; $z$ = 1.00; $p$ = 0.30; condition*group: estimate = 1.24, $z$ = 1.31, $p$ = 0.19). As such, children's ability to initially map the word to the correct referent, regardless of pragmatic demands, was not significantly different between the two age groups.

Next, we examined children's immediate recall performance within and between groups. Both age groups performed significantly above chance across both the Direct Mapping and Inference conditions (chance = 0.25; Figure 2 left). For the Inference condition, younger children had a mean of 0.43 ($CI_{95\%}$ = 0.09) and older children a mean of 0.55 ($CI_{95\%}$ = 0.08). For the Direct Mapping condition, younger children had a mean of 0.38 ($CI_{95\%}$ = 0.08) and older children a mean of 0.57 ($CI_{95\%}$ = 0.08). There was no significant effect of condition for either age group (Younger: $z$ = −1.03, $p$ = 0.30; Older: $z$ = −0.45, $p$ = 0.65). There was a significant effect of initial learning success on immediate recall in the younger children and a similar marginal effect in the older children (Younger: $z$ = 2.13, $p$ = 0.03; Older: $z$ = 1.71, $p$ = 0.09). Comparing the two groups, we found the older children remembered more words than the younger children during immediate recall across both conditions ($z$ = 2.67, $p$ = 0.008). There was no significant interaction between condition and age group ($z$ = 0.33, $p$ = 0.74). Further model parameters can be found in Table 2.

Finally, we examined children's retention performance within and between groups. Both age groups performed significantly above chance across both the Direct Mapping and Inference conditions (chance = 0.25; Figure 2 Right). For the Inference condition, younger children had a mean of 0.35 ($CI_{95\%}$ = 0.08) and older children a mean of 0.52 ($CI_{95\%}$ = 0.09). For the Direct Mapping condition, younger children had a mean of 0.41 ($CI_{95\%}$ = 0.09) and older children a mean of 0.42 ($CI_{95\%}$ = 0.09). Only the older group showed a significant condition effect ($z$ = −2.39,

**TABLE 2** | Model parameters for younger vs. older children for immediate recall and retention.

| Effects | Immediate recall | | | Retention | | |
|---|---|---|---|---|---|---|
| | Estimate | SE | $z$ | Estimate | SE | $z$ |
| Younger children | | | | | | |
| Intercept | −0.66 | 0.23 | −2.86 ** | −0.79 | 0.27 | −2.95 ** |
| Condition (inference vs. direct mapping) | −0.47 | 0.46 | −1.03 | 0.32 | 0.39 | 0.83 |
| Previously learned | 0.89 | 0.42 | 2.13 * | 0.52 | 0.47 | 1.38 |
| Older children | | | | | | |
| Intercept | 0.16 | 0.26 | 0.61 | −0.35 | 0.18 | −1.99 * |
| Condition (inference vs. direct mapping) | −0.22 | 0.49 | −0.45 | −0.84 | 0.35 | −2.39 * |
| Previously learned | 0.79 | 0.46 | 1.71 | 0.92 | 0.38 | 2.44 * |

*$p < 0.1$, **$p < 0.05$, ***$p < 0.01$, **** $p < 0.001$.

$p$ = 0.02), where inferred words were better retained than the directly mapped words, but no significant difference emerged between conditions for younger children ($z$ = 0.83, $p$ = 0.41). Interestingly, the initial mapping accuracy was only a significant factor for retention for older children ($z$ = 2.44, $p$ = 0.01) but not for younger children ($z$ = 1.38, $p$ = 0.17) across conditions. Our combined model yielded a marginal effect of age group ($z$ = 1.75, $p$ = 0.08) and critically a significant interaction between condition and age group ($z$ = 2.05, $p$ = 0.04). Further model parameters can be found in Table 2 below.

Our supplementary analysis revealed a significant interaction of testing phase, condition, and age group ($z$ = −2.06, $p$ = 0.04), driven primarily by the significant effect of testing phase in older but not younger children (older: $z$ = −2.22, $p$ = 0.03. younger: $z$ = 1.62, $p$ = 0.11. See Supporting Table A1). Thus, the younger children recalled fewer words overall but had stable memory for those words, while older children had better initial retention but were only able to sustain this advantage for pragmatically inferred words, with those directly mapped words falling to the rates of their younger counterparts.

**2.5.1.2 | Exploratory Eye-Tracking Analysis.** To better understand the advantage seen for pragmatic inference during retention, we turned back to immediate mappings to examine the time course of eye-gaze in the window ranging from −200 ms before to 1800 ms after the onset of the novel word—the final, disambiguating word in the sentence.

This step allowed us to refine our interpretation of the behavioral outcomes bearing on whether children can use pragmatics to infer word meanings by looking at the relative ease or difficulty of inference resolution for each age group. For instance, if the two groups showed similar rates of pragmatic inference resolution (as is what occurred), we were interested in whether the time course data would also be similar, indicating comparable processing speed in integrating and selecting the referent, or different, such that one group would have greater difficulty in processing and integrating pragmatic cues. As such, the aim we were seeking to better refine with this analysis was not specifically retention but whether children can use pragmatics to infer word meanings and how this may differ between age groups.

Looking at the full analysis window, we found that for both groups, there were significantly more looks to the target than competitor ($ps$ < 0.001), but there was a significant interaction between group and looks-toward-target versus competitor, such that the older children fixated on the target for a greater duration of the time ($\text{Mean}_{\text{older}}$ = 64%, $\text{Mean}_{\text{younger}}$ = 59%, $\beta$ = −0.21, SE = 0.08, $z$ = −2.52, $p$ = 0.01). This indicates that younger children, despite similar behavioral outcomes, may have shown less robust target fixation than their older counterparts. However, this does not tell us if this occurs because of earlier inference resolution as opposed to lower fixation on the target after differentiation.

Our divergent point analysis found that, while both older and younger children diverged in their looks to the target and competitor prior to the onset of the word, older children did so significantly sooner than their younger counterparts (younger children: Mean = −92.09 ms $\text{CI}_{95\%}$ = 74.91 ms); older children: −188.88 ms, $\text{CI}_{95\%}$ = 120.87 ms; $p$ = 0.003; Figure 3). Note that, as a saccade triggered by the novel word onset—0 ms on our timescale—would not be detected until 300 ms after word onset on average, this prediction is made, in both cases, well in advance. Thus, both older and younger children were able to determine and apply informativity to select the referent before the final word was said. However, older children were able to confidently make this computation more rapidly than their younger counterparts. This greater ease, however small, may in turn support deeper encoding and better retention for word meaning.

An important point to note is that, while the divergence point for the older children was calculated to fall at 188.88 ms, this was less than one bin in, indicating that, in most cases, looks to the target and competitor were significantly different from each other, beginning with the very first bin. As such, it is likely that, as the visual data pattern reflects, the divergence occurred well before the selected window. This is in contrast to the younger children, who continued to diverge at the window's onset.

### 2.5.2 | Analysis of Individual Differences in Word Learning

Next, we examined the impact of our individual difference measures, in particular ToM and age, on word learning. As shown

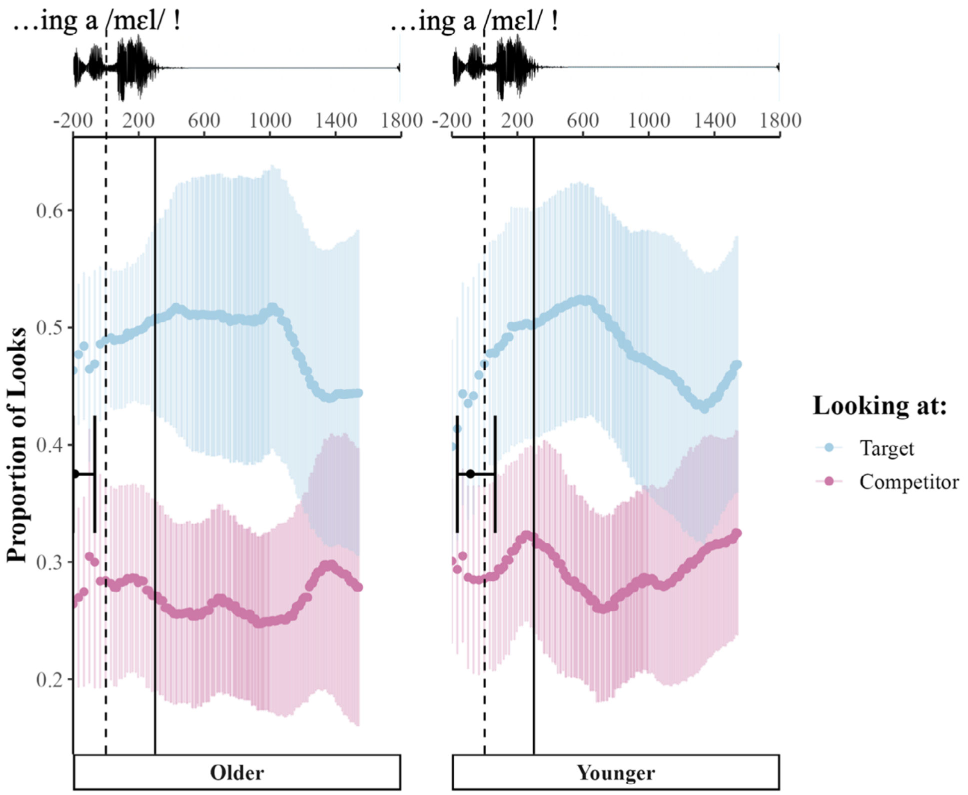


**FIGURE 3** | Divergence of eye-gaze during the novel word for the pragmatic inference condition. The time course across the top is in ms. Solid black points represent divergence points. Error bars represent 95% CI. The vertical dashed line at 0 marks word onset, and the solid vertical line represents the earliest saccade if triggered by word onset. Note that target and competitor proportions do not sum to one; this is because looks away, while not plotted here, are still factored into the relative fixation at any given timepoint.

in Table 1, older children outperformed younger ones on each measure. In subsequent analyses, we treated age as a continuous variable and combined datasets across the older and younger groups. Among the three individual difference measures, only ToM was significantly associated with age (ToM: Pearson's $R = 0.63$, $t = 6.43$, df = 62, $p = 0.00$; Flanker: Pearson's $R = -0.17$, $t = -1.36$, df = 60, $p = 0.18$; Control Question: Pearson's $R = 0.23$, $t = 1.87$, df = 62, $p = 0.07$). Thus, the ToM score improved with age, validating our choice in measure, while the Flanker task and children's accuracy in answering the control questions remained relatively stable across the age range in our sample.

We reported linear regression models for each session—learning, immediate recall, and retention—in each of our two conditions. We found that, unsurprisingly, individuals in the Inference condition that reached higher initial mapping accuracy were also better at retrieving and retaining words (Immediate Recall: $t = 2.86$, $p = 0.01$; Retention: $t = 4.92$, $p = 0.00$). This was less so in Direct Mapping, as only retention accuracy was marginally predicted by initial mapping accuracy ($t = 1.75$, $p = 0.09$). Of our four key individual difference measures, only age significantly and uniquely contributed to the individual variance of pragmatically inferred words ($t = 2.12$, $p = 0.04$, Figure 3). In contrast, age was not a significant predictor of retention for directly mapped words, indicating that this shift was very much specific to inference retention as opposed to a reflection of age on general vocabulary retention ($t = -0.54$, $p = 0.65$, Figure 4). No other variables served as significant predictors for any condition in any session. This included ToM for initial mapping in the Inference condition, suggesting that ToM skills, even though necessary for resolving the pragmatic inference, did not explain individual differences in initial mapping success ($t = 0.16$, $p = 0.88$). Full parameters for the regression models for the Inference condition and for the Direct Mapping condition are available in Supporting Tables A3 and A4, respectively.

Following up on the significant association between age and retention score for the inferred words, three subsequent partial Pearson's correlational analyses were conducted between age and retention score, each controlling for one of the remaining three key covariables (ToM accuracy, Control question accuracy, and Flanker score). Of these three analyses, only the inclusion of ToM accuracy reduced the strength of the relation between age and retention ($r = 0.20$, $t = 1.66$, $p = 0.10$). Age remained as a significant predictor when control question accuracy ($r = 0.34$, $p = 0.01$) or Flanker scores were partialled out ($r = 0.35$, $p = 0.01$). Therefore, we further examined the mediation effect of ToM in the relationship between age and word retention in the Inference condition. We found that ToM mediated the effect of age on Inference condition retention ($F(2, 63) = 3.96$, $p = 0.01$), explaining 11% of the variance, while the direct effect of age after removing the effect of ToM was no longer significant (Figure 5).

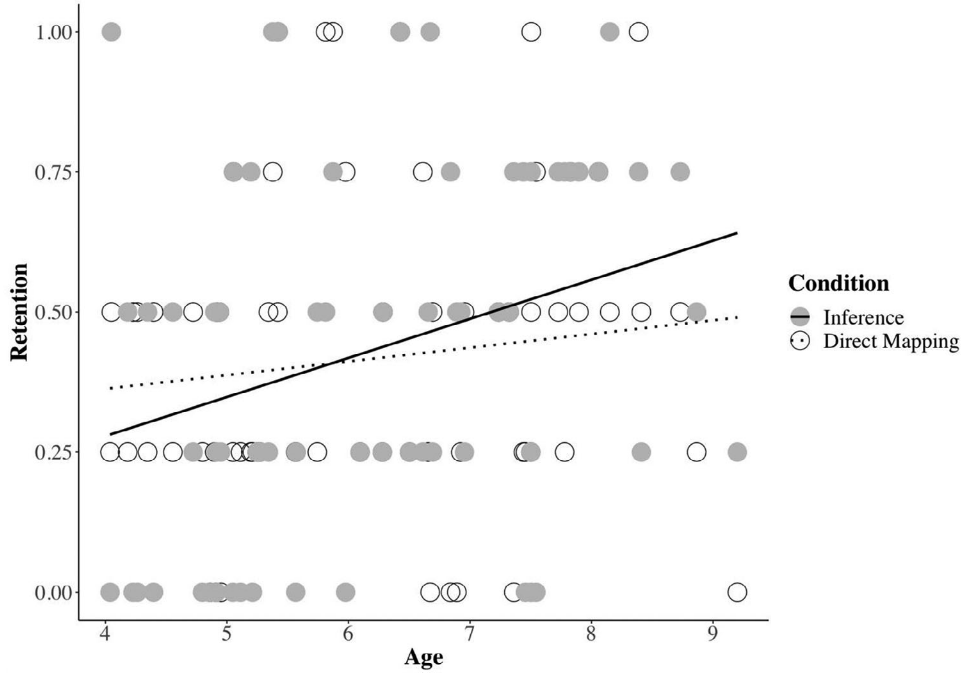


**FIGURE 4** | Age versus retention accuracy by condition. Open circles represent individual direct mapping scores and filled circles Inference scores. The dashed line represents the linear regression line for direct mapping and the solid line the same information for inference. Only the latter is significant.

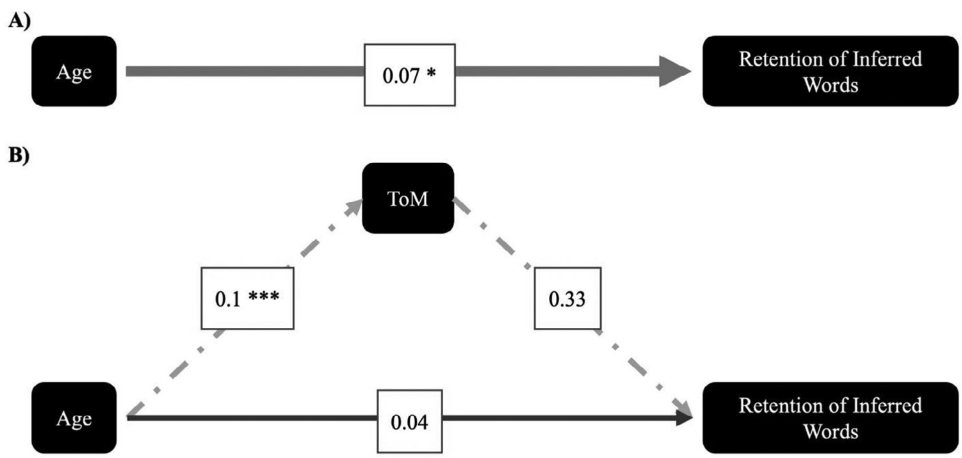


**FIGURE 5** | Mediation of the correlation between age and retention of pragmatically inferred words by ToM. (A) The total effect between age and inferred word retention was significant. (B) The direct effect of age on inferred word retention was no longer significant with ToM removed, while the indirect effect of age on inferred word retention via ToM was significant. Significance: *$p < 0.05$, ***$p < 0.001$.

## 3 | Discussion

### 3.1 | Pragmatic Inference in Word Learning and Memory

The main goal of this work was to determine whether young children can rely on informativeness-based pragmatic inferences to acquire word meanings, and if so, whether they might remember pragmatically inferred words better compared to more directly acquired ones. We found that both 4–6- and 6–9-year-old children are capable of using pragmatic inferences for in-the-moment referential disambiguation of novel words (replicating and extending prior results; Frank and Goodman 2014). Crucially, the ability to do so does not improve with age. This indicates that younger children can reliably apply assumptions of informativeness in pragmatic reasoning and lowers prior estimates of the age at which this reliability emerges, which stands in contrast to much of the literature (Robinson and Whittaker 1985; Feeney et al. 2004). We believe that children's success is due to the nature of the task: instead of relying on overt judgements of informativeness, it involves referential disambiguation: the referent cannot be correctly determined unless a violation of the informativeness maxim is deemed unacceptable (see also Stiller et al. 2015; Kampa and Papafragou 2020, 2023). Our eye-tracking data further confirmed that children below six draw on informativeness just as their older peers: both groups are able to apply assumptions of informativeness significantly before encountering the final word. However, the divergence point between target and competitor comes slightly later in younger children and may indicate a greater degree of hesitancy or effort in using informativity in their pragmatic reasoning. In conjunction with our full window analysis, which shows greater looks-to-target over competitor for older children, we can conclude that, while younger children can reliably apply assumptions of informativity, there are developmental differences in the underlying processing. Such differences may contribute to the lower rate of application of informativity in other studies. It should be noted that other work has found children may become faster at integrating information and making predictions across development (Hartshorne et al. 2015). As such, future work should tease apart whether this divergence is specific to the development of pragmatic inference resolution or more representative of inference resolution across linguistic contexts.

For meaning retention, we found that words learned under both the Inference and Direct Mapping conditions can be retained longer term in both age groups. This indicates that pragmatic inference scenarios (alongside direct mapping) may serve as opportunities for vocabulary growth in the natural environment for both older and younger children. Additionally, we observed robust age effects in both immediate attainment and retention: older children remembered more words than younger children, consistent with the slow developmental course of declarative memory (Ofen et al. 2007).

Most importantly, we found that the age advantage in word retention was particularly evident for the Inference condition. Four-to-6-year-olds did not show a retention advantage of pragmatically inferred word meanings compared to those learned in a less complex scenario. In contrast, 6–9-year-olds retained more words in the Inference condition than the Direct Mapping condition (just like adults in Trice et al. 2023). Thus, pragmatic computation that requires active engagement of cognitive processes during learning does not just serve as an opportunity for vocabulary growth in older children, but in fact enhances it, facilitating memory retention for novel words. This finding extends prior work contrasting word learning scenarios such as Zosh et al. (2013) by bringing it beyond the immediate mapping and into the body of literature of longer-term memory. Unlike Zosh et al. (2013), where word learning in younger children benefited from mutual-exclusivity-based inference, we found a later emerging benefit for a more complex informativity-based pragmatic inference. Therefore, the memory protection effect might require compatible pragmatic difficulty for children's developmental stage. One may predict, for instance, that other pragmatic features such as irony and sarcasm may only boost retention in yet older children who can process them with little effort.

Moreover, our study is the first to report the memory protection effect of pragmatic inference for longer-term retention in children. This sustained memory retention for pragmatically inferred words in older children while those directly mapped continue to be lost echoes much of the literature on longer-term word retention. Words that are mapped in a straightforward, rapid fashion are often forgotten just as quickly in the absence of further memory supports (Axelsson et al. 2012; Horst and Samuelson 2008; Vlach and Sandhofer 2012; Bredemann and Vlach 2021). The pragmatic computation we see here may act as a novel support within the literature through the need to actively engage with the stimuli, similar to requiring pointing in Zosh et al. (2013) or verbally generating tokens in Vlach and Sandhofer (2012). However, unlike such supplementary support mechanisms, pragmatic inferences are a fundamental requirement of resolving the ambiguity and are embedded within the comprehension of the language itself. This offers evidence that linguistic processes and not just extra-linguistic cues or situations may support the sustainable memory of novel words.

An additional consideration in our findings is the role of the novel competitor. Within the experimental design, we chose to use identical syntactic structures for the first sentence of each condition (e.g., "Look! I like this") to ensure that general social and ostensive task engagement motivated by desire for a specific initially-ambiguous object were similar. this lexical setup necessitates identical novel objects in the direct mapping condition, as opposed to a novel competitor being present. As such, the need to compare between distinct competitors is only present in the pragmatic inference condition. Thus, even if the pragmatic inference competitor is not a known object that must be suppressed, its influence on retention outcomes cannot be discounted. However, it is unlikely to be the driving factor of our results for two central reasons. First, we do not see any relationship with other task requiring consideration and suppression of a competitor, be they the control questions or the Flanker task. Second, there is no clear reason why younger children, who benefit from mutual-exclusivity based inferences, would not do so with our more complex pragmatic inferences in this case. In conjunction, these indicates that it is not merely the presence of a competitor that results in better retention. Instead, the complexity of the computation necessitated by the presence of the competitor and the ambiguity it raises seems to play the central role.

### 3.2 | ToM and the Inferential Advantage in Word Learning and Memory

A second major goal of this work was to examine whether ToM skills support more robust memory formation for words acquired inferentially. Indeed, we found that ToM ability mediates the relationship between age and word retention. Importantly, neither the task of executive functioning as an index of general cognitive skill nor the control question accuracy in the ToM booklet task had any significant impact on this relationship. Thus, we conclude that ToM plays a unique and important role in meaning acquisition.

What could the role of ToM be? One possibility is that there is a developing privileging of items processed via ToM in memory. Lin et al. (2019) found that there is a privileged recall ability for socially important information (e.g., faces) over non-social information (e.g., scenes), with the former being more accessible and thereby recalled with greater precision. It is possible that social information in general, and thus information obtained through social cognitive computation that evokes ToM, may have similar privileges.

Another likely possibility is that better ToM skills require fewer resources devoted to the initial computation and thus free up additional resources that can be allotted to active memory encoding. The earlier divergence point in older individuals may act in support of this latter hypothesis, and as such, studies that better tease apart real-time differences in processing should be constructed. However, neither of these possibilities get at the aspect of the issue most central to understanding the role that ToM may be playing in word learning. The question here, then, is not whether ToM plays a role or why it may do so, but at what aspect is its impact most important. Our pragmatic inference resolution can be divided into two key steps: first, the process of computing the speaker's intention and second, the active process of using this computation to select the intended referent and, vitally, to reject a competitor. To tease apart the role that each stage may be playing in facilitating long-term retention, further experimental work would need to be done that splits apart these two aspects from one another. This could be accomplished by either investigating long-term retention via a different type of disambiguation that does not necessitate social cognition, by manipulating the demand of social cognition in the inference condition, or by studying individuals with impaired social communication skills.

Be that as it may, the present results are difficult to reconcile with recent proposals questioning the involvement of ToM in various pragmatic phenomena (Srinivasan et al. 2019; de Marchena et al. 2011; Preissler and Carey 2005; and especially Pijnacker et al. 2009; Andrés-Roqueta and Katsos 2017). In these proposals, the involvement of ToM is neither consistent nor necessary during pragmatic computations. For instance, what might appear to be effects of speaker meaning in computing informativeness might instead arise from a comparison of what the speaker has said to other possible lexical alternatives without actually considering the speaker's intention or mental state. In contrast to this perspective, our data support the position that pragmatic inference involves mentalizing—a position at the heart of classic pragmatic models (Grice 1975; Sperber and Wilson 1986)—and cohere with other evidence pointing to close links between various pragmatic phenomena and ToM abilities in both adults and children (Fairchild and Papafragou 2021; Gollek and Doherty 2016; Matthews et al. 2018; Brosseau-Liard et al. 2015; Sabbagh and Baldwin 2001; among others).

Finally, it is important to consider whether and how ToM is embedded within active encoding that itself facilitates memory (even though no tasks specifically probing active encoding separate from pragmatic inference were included here). Future work could explore the possible correlation between the retention benefits of active encoding environments that have been established in the literature and the retention benefits afforded by pragmatic inference that were discovered here.

### 3.3 | Limitations on Generalizability

When considering the generalizability of these findings, it is important to consider that our ToM task is highly verbal. Although the lack of modulation or correlation effect of children's performance on the control questions on their meaning retention for pragmatic inferences, it is less likely that our results can be explained purely by linguistic or referential disambiguation ability. However, future work should determine if this relationship holds with low and non-verbal ToM tasks to better elucidate ToM independent from general language abilities.

Additionally, while inhibitory control—captured in our Flanker Task—is a salient skill for suppressing a competitor and thus relevant for both the word learning and ToM task, other aspects of EF, such as attention and working memory are likely relevant and uncontrolled (Ackermann et al. 2020). Future work should systematically examine the role of different EF components and tasks in word learning. In addition, while the greater congruency effect in younger than older children is numerically in the expected direction, a lack of significant age-related changes should be considered when choosing a proper EF measure in future studies (Rueda et al. 2004; Massonnié et al. 2019).

Our experimental design introduces several asymmetries between the direct-mapping and pragmatic inference conditions that may influence subsequent word retention. First, the pragmatic inference condition has one additional novel object on the display as a competitor, which may elicit interest and increase children's attention. Although children's abilities to suppress competitors (measured by Flanker and ToM control questions) were not related to their success in retaining words in the pragmatic inference condition, null results are not conclusive. Thus, a comparison condition that involves a novel competitor but does not require pragmatic inference in future studies will help us tease apart the role of novelty from the role of pragmatic computation.

Second, the novel noun was preceded by a definite determiner *this* in the direct mapping condition, but preceded by an indefinite determiner *a* in the pragmatic inference condition. The definite determiner in "Look! I like this [novel object]" means that the novel object itself is what should be selected, which may lead to increased saliency and attention to the novel object in the direct mapping condition than that in the pragmatic condition. This asymmetry of our experimental design should have favored meaning retention in direct mapping. Our findings, however, suggest the opposite pattern, further supporting the robustness of the effect of pragmatic inference on meaning retention. Thus, while this limitation did not seem to significantly influence our conclusions, future work should further investigate the impact of determiners on processing and word learning.

Finally, it should be noted that the sample may not have the racial diversity necessary for full generalizability, especially as both the ToM task and word learning have social components that may interact with perceived group membership. Thus, future work replicating this in a more diverse sample is necessary.

### 3.4 | Implications and Extensions

Beyond their theoretical significance, the present data have real-world implications. If inferential encoding in word learning, in general, and use of ToM, in particular, boost retention for children of school age and above, constructing scenarios that evoke these conditions inside and outside the classroom could have a significant impact on students' rate of vocabulary growth. This is particularly important for second language learners whose first language may not be the dominant one within their academic or occupational sphere and who often learn their second language through direct instruction as opposed to indirect inference or social interaction. Indeed, social interaction has been found to help phonetic and vocabulary learning in a second language (Kuhl et al. 2003; Jeong et al. 2021; Verga and Kotz 2017). Further work can determine whether this extends to pragmatic inferences and aid in the construction of lessons that better support vocabulary growth.

Furthermore, the present data have the potential to inform the vocabulary development of autistic children. From academic to social to occupational achievement, language skills are the single strongest predictor of both autistic and typically developing individuals' long-term success (Tager-Flusberg et al. 2011). Vocabulary knowledge is a critical part of these skills. However, since ToM affects retention for pragmatically inferred words, autistic children who often struggle with ToM (Baron-Cohen et al. 1985) may not have the same retention benefits that develop with age or may even be unable to successfully compute pragmatic inferences to map newly-heard words onto meanings in the first place. Further work can shed light on how novel word learning and retention via pragmatic inference may differ in this population and how likely it is to encounter such pragmatic inferences in a naturalistic setting.


## Acknowledgments

A special thanks to Angelina DiNardo, Veronica Foster, Kelly Chan, Nikoletta Kennedy, and Sarah Gracia for their assistance in segmenting and coding the eye-tracking videos; Marina Hernandez for her help with initial experiment set-up and data collection; and Leah Heisler for her help with participant recruitment and data collection. Hannah Fox and Krystal Mendez were the voices for the stimuli. Anqi Hu programmed the web based eye tracking segmentation and coding pipeline. We thank childrenhelpingscience.com for their assistance with recruitment. This research was supported by internal funds from the University of Delaware awarded to Zhenghan Qi.


## Ethics Statement

The experiments were undertaken with the understanding and written consent of each participant, with the approval of the appropriate ethics committee at the University of Delaware.

## Conflicts of Interest

The authors declare no conflicts of interest.

## Data Availability Statement

Supplementary materials, including appendixes, further stimuli examples, data sets, and code for data-analysis software can be found at https://osf.io/j5e9p/?view_only=b71f338bdb644c1da6c72137831a8778.

## Endnotes

[1] ToM and control questions can be found in the "ToM_scoring_template" in *Individual Difference Measures Stimuli and Examples* of our OSF page (https://osf.io/j5e9p/?view_only=b71f338bdb644c1da6c72137831a8778) and their context in "ToM Booklet #2 Script" (Question Number$_{\text{ToM}}$: 1–12, 14–17, 19–62, 65. Question Number$_{\text{Control}}$: 13, 63, 64). Note that the original question numbers prior to shortening the task were preserved.

## References


Ackermann, L., R. Hepach, and N. Mani. 2020. "Children Learn Words Easier When They are Interested in the Category to Which the Word Belongs." *Developmental Science* 23, no. 3: e12915. https://doi.org/10.1111/desc.12915.

Andrés-Roqueta, C., and N. Katsos. 2017. "The Contribution of Grammar, Vocabulary and Theory of Mind in Pragmatic Language Competence in Children With Autistic Spectrum Disorders." *Frontiers in Psychology* 8: 996. https://doi.org/10.3389/fpsyg.2017.00996.

Anwyl-Irvine, A. L., J. Massonnié, A. Flitton, N. Kirkham, and J. K. Evershed. 2020. "Gorilla in Our Midst: An Online Behavioral Experiment Builder." *Behavior Research Methods* 52, no. 1: 388–407. https://doi.org/10.3758/s13428-019-01237-x.

Apperly, I. A. 2012. "What Is "Theory of Mind"? Concepts, Cognitive Processes and Individual Differences." *Quarterly Journal of Experimental Psychology* 65, no. 5: 825–839. https://doi.org/10.1080/17470218.2012.676055.

Axelsson, E. L., K. Churchley, and J. S. Horst. 2012. "The Right Thing at the Right Time: Why Ostensive Naming Facilitates Word Learning." *Frontiers in Psychology* 3: 88. https://doi.org/10.3389/fpsyg.2012.00088.

Baayen, R. H. 2008. *Analyzing Linguistic Data: A Practical Introduction to Statistics Using* R. 1st ed. Cambridge University Press.

Baayen, R. H., D. J. Davidson, and D. M. Bates. 2008. "Mixed-Effects Modeling With Crossed Random Effects for Subjects and Items." *Journal of Memory and Language* 59, no. 4: 390–412. https://doi.org/10.1016/j.jml.2007.12.005.

Baillargeon, R., R. M. Scott, and Z. He. 2010. "False-Belief Understanding in Infants." *Trends in Cognitive Sciences* 14, no. 3: 110–118. https://doi.org/10.1016/j.tics.2009.12.006.

Bambini, V., C. Gentili, E. Ricciardi, P. M. Bertinetto, and P. Pietrini. 2011. "Decomposing Metaphor Processing at the Cognitive and Neural Level Through Functional Magnetic Resonance Imaging." *Brain Research Bulletin* 86, no. 3–4: 203–216. https://doi.org/10.1016/j.brainresbull.2011.07.015.

Barner, D., N. Brooks, and A. Bale. 2011. "Accessing the Unsaid: The Role of Scalar Alternatives in Children's Pragmatic Inference." *Cognition* 118, no. 1: 84–93. https://doi.org/10.1016/j.cognition.2010.10.010.

Baron-Cohen, S., T. Jolliffe, C. Mortimore, and M. Robertson. 1997. "Another Advanced Test of Theory of Mind: Evidence From Very High Functioning Adults With Autism or Asperger Syndrome." *Journal of Child Psychology and Psychiatry, and Allied Disciplines* 38, no. 7: 813–822. https://doi.org/10.1111/j.1469-7610.1997.tb01599.x.

Baron-Cohen, S., A. M. Leslie, and U. Frith. 1985. "Does the Autistic Child Have a "Theory of Mind"?" *Cognition* 21, no. 1: 37–46. https://doi.org/10.1016/0010-0277(85)90022-8.

Bašnáková, J., K. Weber, K. M. Petersson, J. van Berkum, and P. Hagoort. 2014. "Beyond the Language Given: The Neural Correlates of Inferring Speaker Meaning." *Cerebral Cortex* 24, no. 10: 2572–2578. https://doi.org/10.1093/cercor/bht112.

Bion, R. A. H., A. Borovsky, and A. Fernald. 2013. "Fast Mapping, Slow Learning: Disambiguation of Novel Word–Object Mappings in Relation to Vocabulary Learning at 18, 24, and 30 Months." *Cognition* 126, no. 1: 39–53. https://doi.org/10.1016/j.cognition.2012.08.008.

Boersma, P., and D. Weenink. 2018. *Praat: Doing Phonetics by Computer*. http://www.praat.org/.

Bredemann, C. A., and H. A. Vlach. 2021. "Where's the Advantage? Mutual Exclusivity Promotes Children's Initial Mapping, but Not Long-Term Memory, for Words Compared to Other Strategies." *Frontiers in Psychology* 12: 686554. https://doi.org/10.3389/fpsyg.2021.686554.

Brosseau-Liard, P., D. Penney, and D. Poulin-Dubois. 2015. "Theory of Mind Selectively Predicts Preschoolers' Knowledge-Based Selective Word Learning." *British Journal of Developmental Psychology* 33, no. 4: 464–475. https://doi.org/10.1111/bjdp.12107.

Carey, S., and E. Bartlett. 1978. "Acquiring a Single New Word." *Proceedings of the Stanford Child Language Conference* 15: 17–29.

Carlson, S. M., and L. J. Moses. 2001. "Individual Differences in Inhibitory Control and Children's Theory of Mind." *Child Development* 72, no. 4: 1032–1053. https://doi.org/10.1111/1467-8624.00333.

Çetinçelik, M., C. F. Rowland, and T. M. Snijders. 2021. "Do the Eyes Have It? A Systematic Review on the Role of Eye Gaze in Infant Language Development." *Frontiers in Psychology* 11: 589096. https://doi.org/10.3389/fpsyg.2020.589096.

R Core Team. 2021. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/.

Coutanche, M. N., and S. L. Thompson-Schill. 2014. "Fast Mapping Rapidly Integrates Information Into Existing Memory Networks." *Journal of Experimental Psychology: General* 143, no. 6: 2296–2303. https://doi.org/10.1037/xge0000020.

Cushman, F., R. Sheketoff, S. Wharton, and S. Carey. 2013. "The Development of Intent-Based Moral Judgment." *Cognition* 127, no. 1: 6–21. https://doi.org/10.1016/j.cognition.2012.11.008.

Diesendruck, G., and L. Markson. 2001. "Children's Avoidance of Lexical Overlap: A Pragmatic Account." *Developmental Psychology* 37, no. 5: 630–641. https://doi.org/10.1037/0012-1649.37.5.630.

Fairchild, S., and A. Papafragou. 2021. "The Role of Executive Function and Theory of Mind in Pragmatic Computations." *Cognitive Science* 45, no. 2: e12938. https://doi.org/10.1111/cogs.12938.

Feeney, A., S. Scrafton, A. Duckworth, and S. J. Handley. 2004. "The Story of Some: Everyday Pragmatic Inference by Children and Adults." *Canadian Journal of Experimental Psychology/Revue Canadienne De Psychologie Expérimentale* 58, no. 2: 121–132. https://doi.org/10.1037/h0085792.

Fisher, C., Y. Gertner, R. M. Scott, and S. Yuan. 2010. "Syntactic Bootstrapping." *WIREs Cognitive Science* 1, no. 2: 143–149. https://doi.org/10.1002/wcs.17.

Frank, M. C., and N. D. Goodman. 2014. "Inferring Word Meanings by Assuming That Speakers are Informative." *Cognitive Psychology* 75: 80–96. https://doi.org/10.1016/j.cogpsych.2014.08.002.

Geurts, B., M. Kissine, and B. van Tiel. 2019. "Pragmatic Reasoning in Autism." In *Thinking, Reasoning, and Decision Making in Autism*, edited by K. Moranyi and R. Byrne, 113–134. Routledge.

Gleitman, L. 1990. "The Structural Sources of Verb Meanings." *Language Acquisition* 1, no. 1: 3–55. https://doi.org/10.1207/s15327817la0101_2.

Gollek, C., and M. J. Doherty. 2016. "Metacognitive Developments in Word Learning: Mutual Exclusivity and Theory of Mind." *Journal of Experimental Child Psychology* 148: 51–69. https://doi.org/10.1016/j.jecp.2016.03.007.

Grice, H. P. 1975. "Logic and conversation." In *Syntax and Semantics: Speech Acts*, edited by R. Cole and J. Morgan, Academic Press.

Grigoroglou, M., and A. Papafragou. 2019a. "Interactive Contexts Increase Informativeness in Children's Referential Communication." *Developmental Psychology* 55, no. 5: 951–966. https://doi.org/10.1037/dev0000693.

Grigoroglou, M., and A. Papafragou. 2019b. "Children's (and Adults') Production Adjustments to Generic and Particular Listener Needs." *Cognitive Science* 43, no. 10: e12790. https://doi.org/10.1111/cogs.12790.

Gweon, H., D. Dodell-Feder, M. Bedny, and R. Saxe. 2012. "Theory of Mind Performance in Children Correlates With Functional Specialization of a Brain Region for Thinking About Thoughts." *Child Development* 83, no. 6: 1853–1868. https://doi.org/10.1111/j.1467-8624.2012.01829.x.

Halberda, J. 2003. "The Development of a Word-Learning Strategy." *Cognition* 87, no. 1: B23–B34. https://doi.org/10.1016/S0010-0277(02)00186-5.

Halberda, J. 2006. "Is This a Dax Which I See Before Me? Use of the Logical Argument Disjunctive Syllogism Supports Word-Learning in Children and Adults." *Cognitive Psychology* 53, no. 4: 310–344. https://doi.org/10.1016/j.cogpsych.2006.04.003.

Hartley, C., L.-A. Bird, and P. Monaghan. 2020. "Comparing Cross-Situational Word Learning, Retention, and Generalisation in Children With Autism and Typical Development." *Cognition* 200: 104265. https://doi.org/10.1016/j.cognition.2020.104265.

Hartshorne, J. K., R. Nappa, and J. Snedeker. 2015. "Development of the First-Mention Bias." *Journal of Child Language* 42, no. 2: 423–446. https://doi.org/10.1017/S0305000914000075.

Hornick, S., and E. Shetreet. 2022. "Pragmatic Inferences: Neuroimaging of Ad-Hoc Implicatures." *Journal of Neurolinguistics* 64: 101090. https://doi.org/10.1016/j.jneuroling.2022.101090.

Horst, J. S., and L. K. Samuelson. 2008. "Fast Mapping but Poor Retention by 24-Month-Old Infants." *Infancy* 13, no. 2: 128–157. https://doi.org/10.1080/15250000701795598.

Ito, A., and P. Knoeferle. 2022. "Analysing Data From the Psycholinguistic Visual-World Paradigm: Comparison of Different Analysis Methods." *Behavior Research Methods* 55: 3461–3493. https://doi.org/10.3758/s13428-022-01969-3.

Jeong, J., E. E. Franchett, C. V. Ramos de Oliveira, K. Rehmani, and A. K. Yousafzai. 2021. "Parenting Interventions to Promote Early Child Development in the First Three Years of Life: A Global Systematic Review and Meta-Analysis." *PLoS Medicine* 18, no. 5: e1003602. https://doi.org/10.1371/journal.pmed.1003602.

Kampa, A., and A. Papafragou. 2020. "Four-Year-Olds Incorporate Speaker Knowledge Into Pragmatic Inferences." *Developmental Science* 23: e12920. https://doi.org/10.1111/desc.12920.

Kampa, A., and A. Papafragou. 2023. "Children and Adults Apply Pragmatic Principles When Interpreting Non-Linguistic Symbols." *Journal of Memory and Language* 131: 104429. https://doi.org/10.1016/j.jml.2023.104429.

Kandel, M., and J. Snedeker. 2024. "Assessing Two Methods of Webcam-Based Eye-Tracking for Child Language Research." *Journal of Child Language* 1–34. https://doi.org/10.1017/S0305000924000175.

Katsos, N., and D. V. Bishop. 2011. "Pragmatic Tolerance: Implications for the Acquisition of Informativeness and Implicature." *Cognition* 120, no. 1: 67–81. https://doi.org/10.1016/j.cognition.2011.02.015.

Krigolson, O. E., D. Cheng, and G. Binsted. 2015. "The Role of Visual Processing in Motor Learning and Control: Insights From Electroencephalography." *Vision Research* 110:(Pt B), 277–285. https://doi.org/10.1016/j.visres.2014.12.024.

Kucker, S. C., and L. K. Samuelson. 2012. "The First Slow Step: Differential Effects of Object and Word-Form Familiarization on Retention of Fast-Mapped Words." *Infancy* 17, no. 3: 295–323. https://doi.org/10.1111/j.1532-7078.2011.00081.x.

Kuhl, P. K., F. M. Tsao, and H. M. Liu. 2003. "Foreign-Language Experience in Infancy: Effects of Short-Term Exposure and Social Interaction on Phonetic Learning." *Proceedings of the National Academy of Sciences of the United States of America* 100, no. 15: 9096–9101. https://doi.org/10.1073/pnas.1532872100.

Li, P., and H. Jeong. 2020. "The Social Brain of Language: Grounding Second Language Learning in Social Interaction." *NPJ Science of Learning* 5: 8. https://doi.org/10.1038/s41539-020-0068-7.

Lin, H., W. P. Li, and S. Carlson. 2019. "A Privileged Working Memory state and Potential Top-Down Modulation for Faces, Not Scenes." *Frontiers in Human Neuroscience* 13: 2. https://doi.org/10.3389/fnhum.2019.00002.

de Marchena, A., I. M. Eigsti, A. Worek, K. E. Ono, and J. Snedeker. 2011. "Mutual Exclusivity in Autism Spectrum Disorders: Testing the Pragmatic Hypothesis." *Cognition* 119, no. 1: 96–113. https://doi.org/10.1016/j.cognition.2010.12.011.

Markman, E. M., and G. F. Wachtel. 1988. "Children's Use of Mutual Exclusivity to Constrain the Meanings of Words." *Cognitive Psychology* 20, no. 2: 121–157.

Massonnié, J., C. J. Rogers, D. Mareschal, and N. Z. Kirkham. 2019. "Is Classroom Noise Always Bad for Children? The Contribution of Age and Selective Attention to Creative Performance in Noise." *Frontiers in Psychology* 10: 381. https://doi.org/10.3389/fpsyg.2019.00381.

Matthews, D., H. Biney, and K. Abbot-Smith. 2018. "Individual Differences in Children's Pragmatic Ability: A Review of Associations With Formal Language, Social Cognition, and Executive Functions." *Language Learning and Development* 14, no. 3: 186–223. https://doi.org/10.1080/15475441.2018.1455584.

Nadig, A. S., and J. C. Sedivy. 2002. "Evidence of Perspective-Taking Constraints in Children's On-Line Reference Resolution." *Psychological Science* 13, no. 4: 329–336. https://doi.org/10.1111/j.0956-7976.2002.00460.x.

Naigles, L. 1990. "Children Use Syntax to Learn Verb Meanings." *Journal of Child Language* 17, no. 2: 357–374. https://doi.org/10.1017/S0305000900013817.

Ofen, N., Y. C. Kao, P. Sokol-Hessner, H. Kim, S. Whitfield-Gabrieli, and J. D. Gabrieli. 2007. "Development of the Declarative Memory System in the Human Brain." *Nature Neuroscience* 10, no. 9: 1198–1205. https://doi.org/10.1038/nn1950.

Ovans, Z., A. Ayala, R. Asmah, et al. 2023. "The Feasibility of Online (Virtual-World) Eye-Tracking With Young Children." *Annual Human Sentence Processing Conference*, Pittsburgh, PA.

Papafragou, A., and M. Grigoroglou. 2021. "Contributions of Pragmatics to Word Learning and Interpretation." In *The Oxford Handbook of the Mental Lexicon*, edited by A. Papafragou, J. C. Trueswell, and L. R. Gleitman, Oxford University Press.

Pijnacker, J., P. Hagoort, J. Buitelaar, J.-P. Teunisse, and B. Geurts. 2009. "Pragmatic Inferences in High-Functioning Adults With Autism and Asperger Syndrome." *Journal of Autism and Developmental Disorders* 39, no. 4: 607–618. https://doi.org/10.1007/s10803-008-0661-8.

Pino Escobar, G., A. Tuninetti, M. Antoniou, and P. Escudero. 2023. "Understanding Preschoolers' Word Learning Success in Different Scenarios: Disambiguation Meets Statistical Learning and eBook Reading." *Frontiers in Psychology* 14: 1118142. https://doi.org/10.3389/fpsyg.2023.1118142.

Preissler, M. A., and S. Carey. 2005. "The Role of Inferences About Referential Intent in Word Learning: Evidence From Autism." *Cognition* 97, no. 1: B13–B23. https://doi.org/10.1016/j.cognition.2005.01.008.

Qi, Z., J. Love, C. L. Fisher, and S. Brown-Schmidt. 2020. "Referential Context and Executive Functioning Influence Children's Resolution of Syntactic Ambiguity." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 46, no. 10: 1922–1947. https://doi.org/10.1037/xlm0000886.

Radvansky, G. A., A. C. Doolen, K. A. Pettijohn, and M. Ritchey. 2022. "A New Look at Memory Retention and Forgetting." *Journal of Experimental Psychology. Learning, Memory, and Cognition* 48, no. 11: 1698–1723. https://doi.org/10.1037/xlm0001110.

Revelle, W. 2023. *Psych: Procedures for Personality and Psychological Research*. Northwestern University, R package version 2.3.6. https://CRAN.R-project.org/package=psych.

Richardson, H., G. Lisandrelli, A. Riobueno-Naylor, and R Saxe. 2018. "Development of the Social Brain From Age Three to Twelve Years." *Nature Communications* 9: 1027. https://doi.org/10.1038/s41467-018-03399-2.

Robinson, E. J., and S. J. Whittaker. 1985. "Children's Responses to Ambiguous Messages and Their Understanding of Ambiguity." *Developmental Psychology* 21, no. 3: 446–454. https://doi.org/10.1037/0012-1649.21.3.446.

Rueda, M. R., J. Fan, B. D. McCandliss, et al. 2004. "Development of Attentional Networks in Childhood." *Neuropsychologia* 42, no. 8: 1029–1040. https://doi.org/10.1016/j.neuropsychologia.2003.12.012.

Ruggeri, A., D. B. Markant, T. M. Gureckis, M. Bretzke, and F. Xu. 2019. "Memory Enhancements From Active Control of Learning Emerge Across Development." *Cognition* 186: 82–94. https://doi.org/10.1016/j.cognition.2019.01.010.

Sabbagh, M. A., and D. A. Baldwin. 2001. "Learning Words From Knowledgeable Versus Ignorant Speakers: Links Between Preschoolers' Theory of Mind and Semantic Development." *Child Development* 72, no. 4: 1054–1070. https://doi.org/10.1111/1467-8624.00334.

Saxe, R., and N. Kanwisher. 2003. "People Thinking About Thinking People: The Role of the Temporo-Parietal Junction in "Theory of Mind"." *Neuroimage* 19, no. 4: 1835–1842. https://doi.org/10.1016/S1053-8119(03)00230-1.

Scott, R. M., Z. He, R. Baillargeon, and D. Cummins. 2012. "False-Belief Understanding in 2.5-Year-Olds: Evidence From Two Novel Verbal Spontaneous-Response Tasks." *Developmental Science* 15, no. 2: 181–193. https://doi.org/10.1111/j.1467-7687.2011.01103.x.

Setoh, P., R. M. Scott, and R. Baillargeon. 2016. "2.5-Year-Olds Succeed at Traditional False-Belief Tasks With Reduced Executive Demands." *Proceedings of the National Academy of Sciences* 113: 13360–13365.

Sperber, D., and D. Wilson. 1986. *Relevance: Communication and Cognition*. Wiley-Blackwell. 2nd ed. 1995.

Srinivasan, M., R. Foushee, A. Bartnof, and D. Barner. 2019. "Linguistic Conventionality and the Role of Epistemic Reasoning in Children's Mutual Exclusivity Inferences." *Cognition* 189: 193–208. https://doi.org/10.1016/j.cognition.2019.04.001.

Stiller, A., N. D. Goodman, and M. C. Frank. 2015. "Ad-Hoc Implicature in Preschool Children." *Language, Learning, and Development* 11: 176–190. https://doi.org/10.1080/15475441.2014.927328.

Tager-Flusberg, H., L. Edelson, and R. Luyster. 2011. "Language and Communication in Autism Spectrum Disorders." In *Autism Spectrum Disorders*, edited by D. Amaral, G. Dawson, and D. Geschwind, 172–185. Oxford University Press.

Trice, K. M., D. Saratsli, A. Papafragou, and Z. Qi. 2023. "Pragmatics and Social Cognition in Learning and Remembering Words." PsyArXiv, https://doi.org/10.31234/osf.io/gv4an.

Verga, L., and S. A. Kotz. 2017. "Help Me If I Can't: Social Interaction Effects in Adult Contextual Word Learning." *Cognition* 168: 76–90. https://doi.org/10.1016/j.cognition.2017.06.018.

Vlach, H. A., and C. A. DeBrock. 2017. "Remember Dax? Relations Between Children's Cross-Situational Word Learning, Memory, and Language Abilities." *Journal of Memory and Language* 93: 217–230.

Vlach, H. A., and C. A. DeBrock. 2019. "Statistics Learned are Statistics Forgotten: Children's Retention and Retrieval of Cross-Situational Word Learning." *Journal of Experimental Psychology Learning, Memory, and Cognition* 45, no. 4: 700–711. https://doi.org/10.1037/xlm0000611.

Vlach, H. A., and C. M. Sandhofer. 2012. "Fast Mapping Across Time: Memory Processes Support Children's Retention of Learned Words." *Frontiers in Psychology* 3: 46. https://doi.org/10.3389/fpsyg.2012.00046.

Wellman, H. M., D. Cross, and J. Watson. 2001. "Meta-Analysis of Theory-of-Mind Development: The Truth About False Belief." *Child Development* 72: 655–684. https://doi.org/10.1111/1467-8624.00304.

Wellman, H. M., and D. Liu. 2004. "Scaling of Theory-of-Mind Tasks." *Child Development* 75, no. 2: 523–541. https://doi.org/10.1111/j.1467-8624.2004.00691.x.

Woodard, K., L. Pozzan, and J. C. Trueswell. 2016. "Taking Your Own Path: Individual Differences in Executive Function and Language Processing Skills in Child Learners." *Journal of Experimental Child Psychology* 141: 187–209. https://doi.org/10.1016/j.jecp.2015.08.005.

Zosh, J. M., M. Brinster, and J. Halberda. 2013. "Optimal Contrast: Competition Between Two Referents Improves Word Learning." *Applied Developmental Science* 17, no. 1: 20–28. https://doi.org/10.1080/10888691.2013.748420.

## Supporting Information

Additional supporting information can be found online in the Supporting Information section.