Received: 17 February 2022 | Revised: 3 November 2022 | Accepted: 12 December 2022

DOI: 10.1111/desc.13363

Developmental Science WILEY

RESEARCH ARTICLE

Pragmatics aid referent disambiguation and word learning in young children and adults

Natalie Bleijlevens[1,2] | **Friederike Contier**[1,2,3] | **Tanya Behne**[1,2]

[1]Department of Developmental Psychology, University of Göttingen, Göttingen, Germany

[2]Leibniz ScienceCampus "Primate Cognition", Göttingen, Germany

[3]Department of Psychology, University Potsdam, Potsdam, Germany

Correspondence
Natalie Bleijlevens, Department of Developmental Psychology, University of Göttingen, Waldweg 26, 37073 Göttingen, Germany.
Email: natalie.bleijlevens@uni-goettingen.de

Funding information
the Deutsche Forschungsgemeinschaft DFG, Grant/Award Number: 254142454/GRK 2070

Abstract

How do children succeed in learning a word? Research has shown robustly that, in ambiguous labeling situations, young children assume novel labels to refer to unfamiliar rather than familiar objects. However, ongoing debates center on the underlying mechanism: Is this behavior based on lexical constraints, guided by pragmatic reasoning, or simply driven by children's attraction to novelty? Additionally, recent research has questioned whether children's disambiguation leads to long-term learning or rather indicates an attentional shift in the moment of the conversation. Thus, we conducted a pre-registered online study with 2- and 3-year-olds and adults. Participants were presented with unknown objects as potential referents for a novel word. Across conditions, we manipulated whether the only difference between both objects was their relative novelty to the participant or whether, in addition, participants were provided with pragmatic information that indicated which object the speaker referred to. We tested participants' immediate referent selection and their retention after 5 min. Results revealed that when given common ground information both age groups inferred the correct referent with high success and enhanced behavioral certainty. Without this information, object novelty alone did not guide their selection. After 5 min, adults remembered their previous selections above chance in both conditions, while children only showed reliable learning in the pragmatic condition. The pattern of results indicates how pragmatics may aid referent disambiguation and learning in both adults and young children. From early ontogeny on, children's social-cognitive understanding may guide their communicative interactions and support their language acquisition.

KEYWORDS

disambiguation, discourse novelty, mutual exclusivity, pragmatics, retention, word learning

Research Highlights

- We tested how 2-3-year-olds and adults resolve referential ambiguity without any lexical cues.

- In the pragmatic context both age groups disambiguated novel word-object-mappings, while object novelty alone did not guide their referent selection.
- In the pragmatic context, children also showed increased certainty in disambiguation and retained new word-object-mappings over time.
- These findings contribute to the ongoing debate on whether children learn words on the basis of domain-specific constraints, lower-level associative mechanisms, or pragmatic inferences.

1 | INTRODUCTION

Within their first years of life, children become remarkable language users. They acquire the meanings of several novel words per day, seemingly without any effort. Part of a crucial discussion, however, is the underlying mechanisms responsible for their success. Different theoretical accounts have been proposed that can roughly be classified as lexical, pragmatic, and associative accounts of word learning. Lexical constraint accounts explain children's success in word learning by their use of internal lexical biases. These are supposed to limit the hypothesis space of potential referents, for example, by making children focus on whole objects or accept only one category label per object (Golinkoff et al., 1992; Markman & Wachtel, 1988). In contrast to such domain-specific approaches, socio-pragmatic accounts assume word learning to be part of a broader socio-cognitive capacity: Children are assumed to interpret their linguistic environment and infer the meanings of novel words by reasoning about the speaker's intentions (e.g., Tomasello, 2010). And finally, associative accounts provide the most parsimonious proposal regarding children's cognitive equipment for word learning: They claim word learning to be the product of domain-general associative learning mechanisms without the need for any lexical or pragmatic inferences (e.g., Smith, 2000). In this line, associative mechanisms and biases, such as the sensitivity to novelty and perceptual salience, are assumed to be sufficient to explain young children's rapid word-learning progress.

These three theoretical approaches make different assumptions in the interesting test case of *referential ambiguity*: When hearing a novel word (e.g., "car") and without having any prior knowledge to rely on, children's environments offer a huge hypothesis space of potential referents (Quine, 1960). This problem makes word-object mappings complicated on different levels. First, it is unclear to the naïve listener, to which part of an object the word refers: Does "car" refer to the whole vehicle, only to the wheel, to its color or its movement? Second, even if the listener tends to focus on the whole object most of the time (see Markman & Wachtel, 1988), it is still unclear which of the many objects in their visual scene they should focus on. Does "car" refer to the vehicle on the street, the traffic lights, or the road sign? How do children, or any novice in a given language, solve this referential ambiguity? And how can lexical, pragmatic, and associative accounts explain this learning process?

Past research has shown that despite this ambiguity problem, children's referent selection is not arbitrary at all. In many situations, they are able to limit the hypothesis space and identify the intended referent. For instance, if children are presented with one well-known object (e.g., apple) and one novel, unnamed object and hear a novel word (e.g., "modi"), they consistently tend to select or attend to the novel object, a response bias known as the *disambiguation effect* (see, Lewis et al., 2020 for a meta-analysis). This effect has proven to be a very robust phenomenon: It is in place in infants as early as 17 months (Halberda, 2003) or even younger (Pomiechowska et al., 2021) as well as in adults (e.g., Au & Glusman, 1990; Halberda, 2006) and increases in strength with age (Bion et al., 2013; Lewis et al., 2020). However, ongoing debates concern two questions: First, what mechanisms guide disambiguation, and second, does success in referent disambiguation indicate long-term word learning? The current paper aims to touch on both of these issues.

1.1 | Possible mechanisms behind the disambiguation effect

Based on the broader classes of word learning theories, there are three contrasting proposals for possible mechanisms behind the disambiguation effect. First, children's success in disambiguation may be achieved by lexical biases or constraints. The *mutual exclusivity* (ME) bias, for example, states that people assume objects to have only one label and avoid second labels for already named categories (Markman & Wachtel, 1988). Thus, upon hearing "modi" in the face of an apple and a novel object, one would assume: "The apple cannot have two labels. So, 'modi' must refer to the novel object." These constraints (see also Golinkoff et al., 1994; Mervis & Bertrand, 1994) stress the role of lexical knowledge in the disambiguation process. This is supported by empirical work that examined children's task performance in relation to both their overall word knowledge (Bion et al., 2013; Kalashnikova et al., 2016; Mervis & Bertrand, 1994) and their specific word knowledge of objects in a given task (Grassmann et al., 2015; Lewis et al., 2020). However, for young children with their limited vocabulary, it is quite common to be confronted with several potential referents whose labels are unknown, making it hard or impossible to engage in such reasoning by exclusion. Lexical constraint accounts thus cannot explain children's whole success in word learning and it might be reasonable to assume specific lexical knowledge to be helpful, but not essential, for early success in disambiguation.

The second type of proposal suggests that the disambiguation effect is based on pragmatic processes. These socio-pragmatic accounts claim no lexical constraints are needed for disambiguation. Instead, they assume children's referent selection to be based on domain-general socio-cognitive capacities (Diesendruck & Markson, 2001), especially pragmatic inferences about the speaker's intentions, for example, "If you meant the apple, you would have used the conventional word 'apple'. As you said 'modi,' you must be referring to the other thing." This proposal is supported by much research, with both children and adults, showing that listeners take the speaker's perspective, experiences and preferences into account to resolve ambiguity and determine the correct referent of a novel word (e.g., Bohn & Frank, 2019; Clark, 2015). Even 1-year-olds have been found to consider shared experience, recognizing that others refer excitedly to things that are novel from their (not the child's) perspective (e.g., Akhtar et al., 1996; Moll et al., 2006). However, to date it remained unclear whether pragmatic reasoning is necessary for disambiguation or whether even simpler cognitive processes would be sufficient.

Third, there are novelty accounts, offering the most parsimonious explanation of the disambiguation effect. As part of the broader class of associative accounts, they propose that children disambiguate novel words by relying on associative cues and domain-general biases, specifically their attraction to object novelty. Contrary to the accounts above, there is no exclusion process assumed and neither lexical knowledge nor pragmatic inferences are needed to succeed in the disambiguation task. Instead, they claim that it is sufficient to map the novel label to the most novel object in the scene (e.g., Mather & Plunkett, 2012). To date there are mixed findings as to whether the attraction to novelty, without any pragmatic or lexical cues, can be sufficient for disambiguation. Some studies find children to rely on relative object novelty to disambiguate novel words, that is, upon hearing a novel word, children selected/attended to a completely novel object over one that is unknown, but pre-exposed during an earlier experimental phase (Dysart et al., 2016; Horst et al., 2011; Mather & Plunkett, 2012). However, others argue children only consider these cues within the pragmatic context. Children are similarly able to *exclude* the newest object in the scene as the correct referent, if the discourse context suggests this to be sensible (Graham et al., 2005; Grassmann et al., 2009; Marno, 2021) or select objects that are novel for their conversation partner, but not themselves (Akhtar et al., 1996). In this line, previous results in which children favored the most novel object (Dysart et al., 2016; Mather & Plunkett, 2012) could be explained by pragmatic inferences like "If she meant the pre-exposed object, she would have named it earlier."

Horst et al. (2011) argued against a pragmatic interpretation of such results by showing that children favored the "supernovel" object even if the speaker in referent selection trials was different from the speaker present during the pre-exposure of one object. However, it is unclear whether this behavior is due to novelty being the driving mechanism behind disambiguation, or rather due to children's preference for interacting with novel objects in an interactive setting (regardless of any label; see Mather & Plunkett, 2012) and/or their failure to meet the task's demands of tracking the change of speaker (see, e.g., Simons & Rensink, 2005) and registering which of the two speakers had been observing the scene at which stage (Moll et al., 2007; Moll & Tomasello, 2007).

Further, even within studies stressing the role of object novelty in disambiguation results were mixed. For example, in Mather and Plunkett's (2012) study, infants shifted their gaze to a completely novel object (vs. a pre-exposed novel object and a known object) upon hearing a novel label in experiment 1, but the effect could not be replicated in their second experiment (with the presentation of only the novel and pre-exposed object).[1] Especially in the light of the current replication crisis (Frank, Bergelson, et al., 2017), more research is needed to disentangle which source of information or cues are necessary and sufficient for children (and adults) to resolve referential ambiguity.

1.2 | The relation between disambiguation and word learning

Besides the discussion about the mechanisms underlying disambiguation, a more recent debate concerns the relation between children's referent disambiguation and their retention of the novel label. While children seem to show consistent and robust patterns in their referent selection, it is unclear whether this behavior offers insights in how they learn words or only reflects their response to a situational demand: Does the referent selection indicate real long-term word-learning or only an attentional shift in the moment?

Based on earlier research, it has commonly been assumed that children's disambiguation indicates their acquisition of a novel word. After children successfully inferred a certain word-object-mapping, they could remember it after different delays of time from some minutes up to even a week (Carey & Bartlett, 1978; Markson & Bloom, 1997; see also Spiegel & Halberda, 2011; Zosh et al., 2013). However, some recent studies questioned this assumption. Instead, they suggest that children's referent selection is an on-line process that responds to the communicative demands of the moment, but is independent from long-term word learning (McMurray et al., 2012). In fact, Horst and Samuelson (2008) showed that after successful disambiguation, 24-month-olds did not remember the inferred label-object mappings even after a delay of just 5 min (see also Bion et al., 2013). What might explain this vast discrepancy in findings, with early work suggesting children can remember the inferred label-object link even a week later and recent work suggesting that they do not even retain it for 5 min?

The results across studies show that these discrepancies cannot be reduced to a distinct age pattern: While there is some data suggesting that children's retention ability improves with age, with fragile retention even at 30 months (Bion et al., 2013), others already do show retention at this young age (e.g., Spiegel & Halberda, 2011). One possible explanation, however, may be methodological differences; in particular, the number of novel labels and novel objects children were confronted with. Studies showing successful retention (e.g., Carey & Bartlett, 1978; Markson & Bloom, 1997) typically tested the retention of only a single word and might thereby have overestimated children's word-learning. In this case, it is unclear whether children specifically linked a certain word to a certain referent and remembered this link or whether, in the retention task, children simply again selected the only object that had previously been highlighted (see Axelsson & Horst, 2013). Conversely, in some later studies demonstrating children's failure in learning, participants encountered up to eight completely novel words and many more novel objects (e.g., Kucker et al., 2020) which may have overwhelmed them, thereby perhaps underestimating their potential to acquire words based on disambiguation processes. This raises the importance of teaching children more than a single word to test real word learning, while simultaneously preventing too many performance demands (e.g., confronting children with many novel labels and objects within a brief period of time).

1.3 | The current study

To investigate the guiding processes and mechanisms behind both referent disambiguation and word learning, we conducted an online study with 2-3-year-olds as well as adults. We included adults as a comparison group for methodological validation and to assess differences between word learning strategies in children and adults. Additionally, this allowed us to ask adults for verbal explanations of selection strategies.

Our first aim was to shed light onto the mechanisms behind the disambiguation effect. The use of lexical constraints (e.g., the ME bias) is dependent on situations that provide name-known objects in contrast to unfamiliar ones. However, these situations are probably not too common for early word learners who are constantly confronted with several objects which are unknown to them. For this reason, we based our design on situations with both potential referents being unknown. Thus, we controlled for the nameability of objects (thereby also assessing whether lexical knowledge is necessary for disambiguation) and aimed to investigate: Is object novelty sufficient on its own or are additional pragmatic inferences necessary to succeed in disambiguation?

To test this, we designed a disambiguation task in which we confronted participants with an ambiguous labeling event. In each disambiguation trial, we presented a novel label together with two unknown objects that differed in their relative novelty to the participant: one of them pre-exposed and one of them completely novel. We tested whether this ambiguity could be resolved if the only cue given was relative object novelty or whether, in addition, a pragmatic context was needed. In the pragmatic condition, the speaker in the event was present and witnessed when the pre-exposed object appeared, but did not label it, thus this shared background or common ground suggested that the speaker subsequently referred to the novel object. In the non-pragmatic condition, the speaker was absent when the pre-exposed object was introduced, so no pragmatic context was provided. Figure 1 shows the predicted performance in both conditions by different theoretical accounts. We measured participants' object choices and additionally, as a more fine-grained measure, their (un)certainty during the selection. For adults only, we also assessed their verbal explanations of the selection process.

The second aim of this study was to investigate the relation between disambiguation and word learning in our two different learning contexts. Therefore, we tested participants' retention of their selections after a 5-minute-delay in both conditions. We taught each participant two different words, thereby enabling a real word learning test (see Axelsson & Horst, 2013) without overwhelming younger children. Our goal was to examine whether their selection behavior in these different scenarios indicates real word learning leading to the retention of these mappings after a short delay, instead of only an attentional shift in the moment of the selection task. Finally, we wanted to test whether participants' success depends on the cues given for disambiguation: If pragmatic context has a positive effect on word learning, we expected participants to remember the word-object mappings better in the pragmatic than in the non-pragmatic condition.

2 | METHOD

The experimental design, procedure, sample sizes, and statistical analyses were pre-registered on OSF (https://osf.io/p2re6). Study materials, data sets, analysis scripts, as well as details regarding the sample size calculation, counterbalancing plan, and analysis results are accessible here (https://osf.io/bk47x/).

2.1 | Participants

Our final sample consisted of 72 typically developing monolingual 2-3-year-old children (24-48 months, M = 36.5 months, SD = 6.8; 40 female, 32 male) and 112 adults (18–58 years, M = 29.1 years, SD = 8.5; 47 female, 64 male, 1 diverse; 34 multilingual), both with German as their first language. We determined the required number of participants a priori via data simulation. The calculation was based on our main hypotheses testing children's and adults' object choices in referent selection trials in both conditions against chance with the goal to obtain 0.8 power.

In addition to the final sample, 34 children were excluded from analyses based on our pre-registered exclusion criteria and six children provided data for referent selection, but not retention trials. Within the final sample, nine single trials had to be excluded (for details see Appendix A). We recruited children via the department's database of children whose parents previously agreed to the participation in our studies. Adults were recruited via an online recruitment platform (www.prolific.com) to get a variety of adults who had not necessarily had contact to psychological research before, and received £1.95 (=£9.00/h) for their participation.

2.2 | Design

We had a 2 (Condition: "pragmatic" vs. "non-pragmatic") × 2 (Age group: children vs. adults) factorial repeated-measures design. Participants were randomly assigned to one of the two conditions.

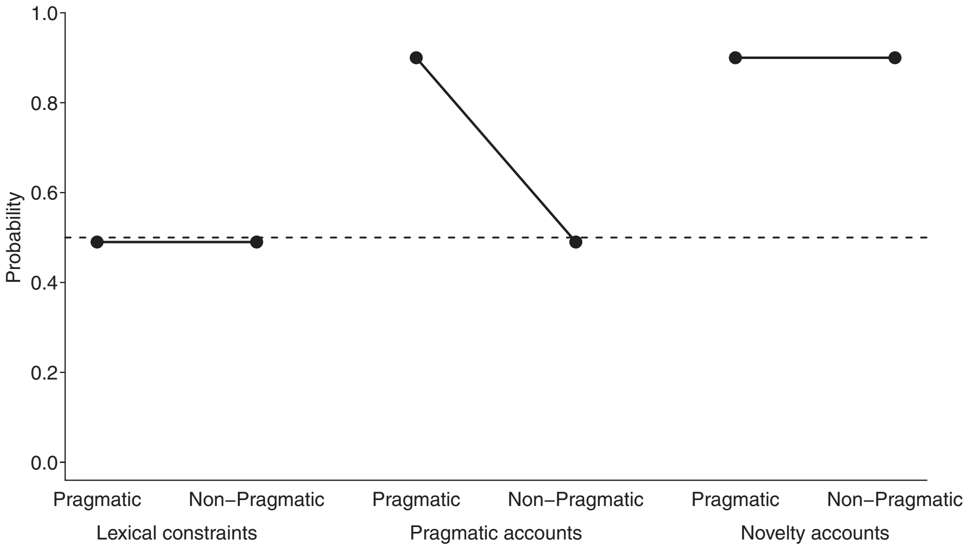

FIGURE 1 Predicted proportion of correct (novel) object choices for both conditions by different theoretical accounts. Dots and solid lines present the predicted patterns of object selections, and the dotted line indicates chance level. If participants rely on lexical cues alone (lexical constraints; left), they should succeed in neither condition, as both do not include known objects which can be excluded. If no lexical information, but pragmatic inferences are needed for disambiguation (pragmatic accounts; middle), they should succeed in the pragmatic condition (selecting the novel object above chance), but not in the non-pragmatic one (not selecting the novel object above chance), additionally revealing a significant difference between both conditions. And only if the attraction by object novelty is sufficient for success in disambiguation (novelty accounts; right), they should select the completely novel object in both conditions above chance. The predicted pattern regarding participants' certainty during selection was identical to the one for their object choices, such that higher proportions of correct object choices should come along with higher certainty in their selection. Only if pragmatic inferences are crucial for disambiguation, participants' certainty should differ between conditions (pragmatic accounts).

2.3 | Stimuli

For visual stimuli, we used images made accessible by Bohn et al. (2022) and open-source material. Pictures of known objects were used in the practice trials and pictures of four unknown objects in the referent selection and retention trials. Auditory stimuli were recorded from three female native speakers of German, one for each speaker in the experimental videos. Two non-words served as novel labels for objects in referent selection ("modi," "toma") and six known words as labels for known objects in practice trials (apple, car, flower, bus, ball, shoe). Each of these German words was produced by at least 70% of the 24-month-old German-speaking children which made up our youngest participants (see Wordbank; Frank et al., 2017). Videos were created via PsychoPy (Peirce et al., 2019) and Powerpoint.

2.4 | Procedure

The experiment was conducted online as an unmoderated remote study (e.g., Rhodes et al., 2020) via Labvanced (Goeke et al., 2017) due to the Covid-19 pandemic. Participants joined the study from their own computers or laptops at home. Participation via smaller devices (smartphones/tablets) was not permitted. Children took part with their parents who were informed about the study via videos and instruction texts. For children (but not adults), the session was video-recorded. Via these recordings, we ensured the correspondence of parents' clicks and children's actual pointing gestures (see below) and the exclusion of trials in which children were distracted, parents interfered, or technical issues occurred.

After giving informed consent, each participant took part in four phases: practice-I (4 trials), practice-II (2 trials), referent selection (2 trials), and retention (2 trials). On each trial an animal character (located at center stage and looking straight ahead) asked for a referent, with the objects to choose from presented on a row of tables below (Figure 2). If children had not responded after 11 s, parents were asked to encourage their children to answer by reading out "What do you think is the right one? Just take what you think is right." Adult participants were asked about their selection strategy at the very end of the study. See Appendix B for counterbalancing information. The full transcript, randomization plan and video examples are uploaded on OSF (https://osf.io/bk47x/).

2.4.1 | Practice-I

The first speaker (frog) introduced herself. On each of the four trials, two known objects appeared at the top of the screen, one above each table, and descended until each rested on that table. Frog was standing behind them and asked participants to show her the object she was labeling: "Oh! There is a [known label]! Look at the [known label]! Can you show me the [known label]?". Participants' reaction time was unlimited. Parents were asked to help their children until they provided the correct answer and children received positive feedback.

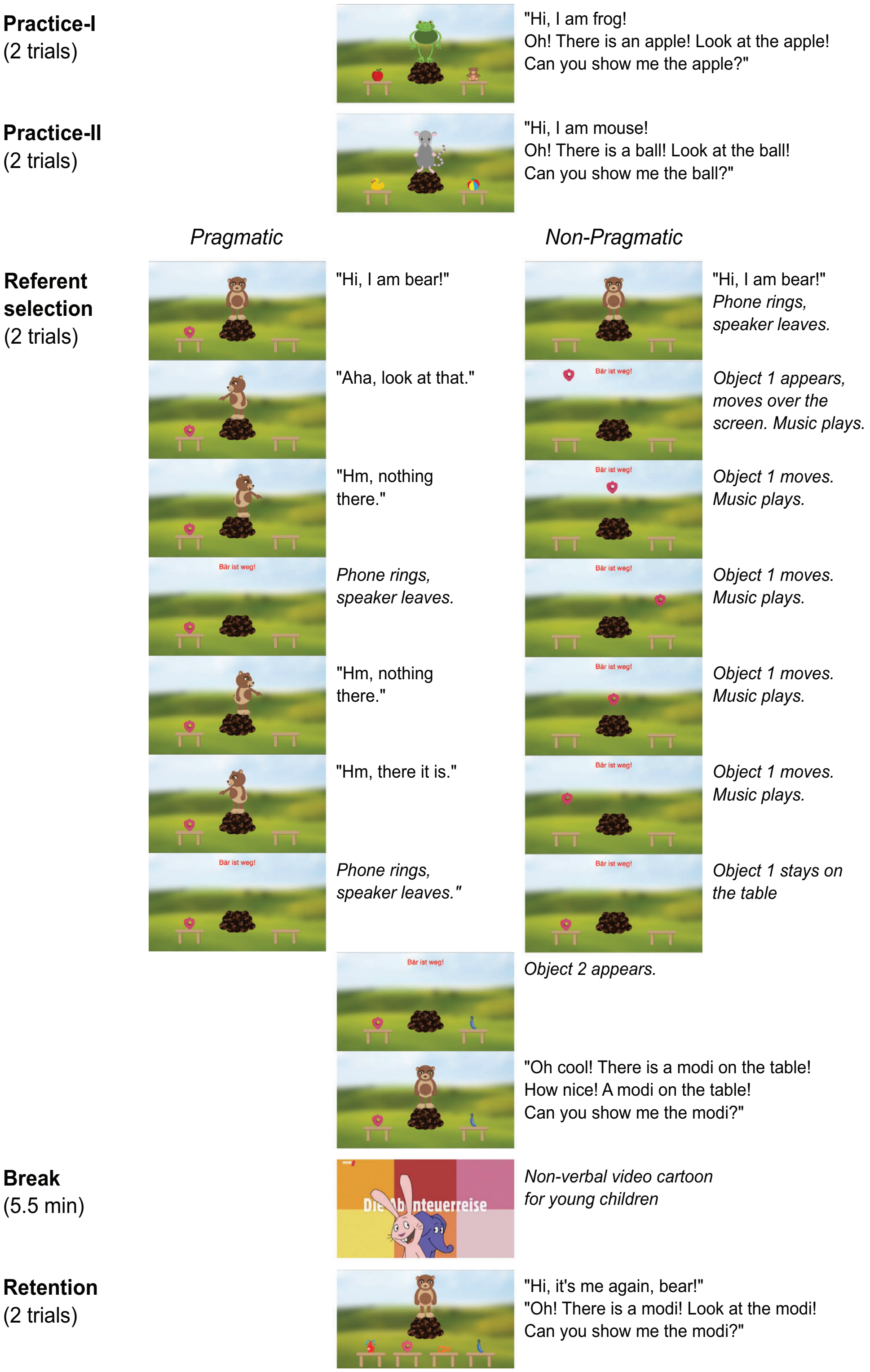

FIGURE 2 Experimental procedure: Example trials for each phase (translated from German original).

2.4.2 | Practice-II

The subsequent two practice trials followed the same general procedure as the first four, but differed in the following respects: (a) the response time was limited to 33 s after the speaker (mouse) spoke her last word, (b) parents were instructed to not help their children anymore, (c) children did not receive any feedback after their choice, and (d) trial order and target location were randomized (see Appendix B).

2.4.3 | Referent selection

The two referent selection trials started in presence of the speaker (bear) who introduced herself. Afterward, the first unknown object was pre-exposed in one of two ways (i.e., in presence or absence of bear), depending on condition. Whenever bear left the scene or appeared again, we presented text on the screen to emphasize her absence/presence: "Bear is gone!"/"Bear is back." Parents were asked to read this out for their children. The video presentation time of each object was matched across conditions. In both conditions, the pre-exposed object was presented for 32 s, before the novel object appeared, with both then shown for 16 s. After each scene with a novel event, parents/adults clicked on a button to continue and, therefore, the exact object exposure times depended on how fast they clicked in between the scenes.[2] See Appendix B for counterbalancing information.

Pragmatic condition. The first unknown object (*pre-exposed object*) was shown laying on one of the tables; the other table was empty (Figure 2). Bear either first registered the presence of the object by looking and pointing to it and saying "Aha, look at that", and then looked and pointed to the empty table and said "Hmm, nothing there" or vice versa. Then, the phone rang and bear left the scene by jumping inside the small hill. After her return, she again commented on the empty table ("Hmm, nothing there") and the pre-exposed object ("Hmm, there it is") in the opposite order than before. The phone rang again and bear left for a second time.

From here on, both conditions were identical: In bear's absence, the second unknown object (*novel object*) fell down on the empty table. Bear reappeared and, without changing her frontal gaze direction, she said excitedly: "Oh cool, there is a [novel label] on the table! How nice! A [novel label] on the table! Can you show me the [novel label]?" The correct choice was the novel object. Participants' response time was limited to 33 s after bear finished speaking. This condition was based on the procedure in Bohn et al.'s (2022) second study (condition "same speaker").

Non-pragmatic condition. Both tables were empty at the beginning. The phone rang and bear left the scene by jumping inside the small hill. In her absence, the first unknown object (*pre-exposed object*) appeared and moved over the screen for the same time as it would have been laying on the table in the pragmatic condition. Background music was playing to ensure equal attention to the scene as in the pragmatic condition.[3] The object stayed stationary on the table from the point in time in which bear would have left for the second time in the pragmatic condition. From here on, both conditions were identical (see above). The "correct" choice was again the novel object.

2.4.4 | Break

A break of 5.5 min served as time delay prior to retention trials. Participants watched an animated and non-verbal video cartoon for young children that was unrelated to the study materials.

2.4.5 | Retention

The speaker from referent selection trials (bear) was standing behind four tables. Four objects fell down on these tables: the two target objects and the two distractor objects from both referent selection trials. We presented all four objects again to have a more stringent retention test: Participants needed to remember specific label-object links (not just which of two objects they had chosen previously). Presenting all four objects also allowed us to assess retention in all participants (not only in those participants who had selected the object *we* defined as the target) and the setup did not allow for cross-situational learning, so that retention itself could be assessed. On each of the two retention trials, bear labeled one of the previous targets from referent selection trials and said: "Oh, there is a [novel label]! Look at the [novel label]! Can you show me the [novel label]?". Participants' response times were again limited to 33 s. See Appendix B for counterbalancing information.

2.5 | Measures

2.5.1 | Object choices

In referent selection trials, the selection of the novel (vs. pre-exposed) object counted as the correct choice in both conditions. In retention trials, a consistent choice was defined as the selection of the same object that had previously been selected in the corresponding referent selection trial. We analyzed consistent instead of correct choices in retention trials, because participants defined the targets in our ambiguous labeling events themselves (as they were not given any feedback on their referent selection). We recorded object choices via adults' mouse clicks (adult version) or parents' confirming mouse clicks after children's pointing response (child version). For a random sample of webcam recordings, a blinded coder confirmed that, in the codable videos, parents' clicks always corresponded to children's pointing gestures.[4]

2.5.2 | Uncertainty

For adults and children, we measured response times, that is, the time before either adult clicked (adult version) or children finished pointing, as indicated by the video recording (child version). We started the coding of children's response times with label onset of the first target mention.[5] For children, we also measured social referencing in a binary manner, that is, at least one (vs. no) look to the parent before or during the child's response. Children's video recordings were coded by a trained, blinded coder. Unclear cases of social referencing (2 out of

TABLE 1 Categories of adults' selection strategies

Strategy	Explanation	Examples
Speaker intent	Reasoning based on the speaker, her behavior or intentions	"Bear was excited about the new object" "Bear didn't name the first object. Only when the second object appeared he named it"
Perceptual features	Reasoning based on perceptual (e.g., visual or auditory) properties and salience of the objects	"I inferred the shape from the sound of the word" "The flying object was visible for a longer time and got more attention"
Experimental logic	Reasoning based on the pragmatics of the experiment/the intentions of the experimenters	"I thought I was meant to concentrate on the object that was flying around" "The object was probably shown for longer to learn it better"
Explicit guessing	Indication of own ignorance/selection based on intuition	"I relied on my gut feeling" "I just guessed"

148) were discussed between two coders. Reliability coding by another blinded coder for 32 recordings (>20%) revealed 100% agreement.

2.5.3 | Adults' selection strategies

In an exploratory manner, we coded adults' open explanations of their selection strategies. We created four categories: "speaker intent," "perceptual features," "experimental logic," and "explicit guessing" (Table 1). Whenever participants indicated more than one strategy, we coded the strategy that got more emphasis or, if not applicable, the one related to the participant's first trial. Unclear cases (8 out of 112) were discussed between two coders. Reliability coding by another blinded coder for 24 explanations (>20%) revealed 88% agreement.

2.6 | Statistical analyses

2.6.1 | Confirmatory analyses

We analyzed the data using R (version 4.0.3; R Core Team, 2020), following our preregistered analysis plan.[6] See Appendix C for the packages and functions used. The data set, descriptive data, R scripts, detailed model results and assumption tests are accessible on OSF (https://osf.io/bk47x/).

For each model with more than one predictor, we used Likelihood Ratio Tests to compare the fit of the full model to that of a null model which was identical but lacked the predictors of interest. This way, we tested for the overall effect of our fixed effects and avoided "cryptic multiple testing" (Forstmeier & Schielzeth, 2011). If not stated otherwise, the model assumptions were met.

To test whether participants' performance differed from chance on their referent selection trials, we fitted a binomial mixed effects model. We included "group" (i.e., each combination of condition and age group) as predictor, suppressed the intercept,[7] and added random intercepts for participants. This model tested performance against 0.5 (chance level), yielding separate values for each combination of age group and condition.

To test for differences between conditions and age groups, we conducted another binomial mixed effects model. We predicted correct choices in referent selection trials by condition, age group and their interaction and added random intercepts for participants.[8] In addition to the predictions detailed above (Figure 1), we expected adults to perform better than children in the pragmatic condition, and expected the reverse age pattern in the non-pragmatic condition.

Regarding participants' (un)certainty, we first tested whether response times on referent selection trials differed between conditions. Therefore, we conducted two separate linear mixed effects models—one analyzing children's response times and one adults'[9]—including condition as a predictor and random intercepts for participants. In both models, the assumptions of normality and homogeneity of the residuals were violated. Therefore, and because it is advisable for this kind of data, we log-transformed response times. Second, we compared children's social referencing in both conditions: We predicted social referencing by condition in a binomial mixed effects model and included random intercepts for participants. For children's uncertainty analyses, we adapted the alpha level to 0.025 due to multiple testing: uncertainty measured as both response times and social referencing.

To test whether participants remembered their previous selections after a short delay, we ran a binomial mixed effects model on participants' consistent choices in retention trials. We included "group" as predictor, suppressed the intercept and added random intercepts for participants. We then tested performance against chance level of 0.25, separately for each group.

To compare the retention performance between conditions and age groups, we ran a binomial mixed effects model on participants' consistent choices. We included condition, age group, and their interaction as predictors and added random intercepts for participants. Besides the expectations described above, we expected adults to show overall higher retention abilities than children in both conditions.

2.6.2 | Exploratory analyses

In an exploratory manner, we tested for (a) differences in adults' strategies across conditions, (b) age effects in children's retention performance and whether their object choices in referent selection trials affected their retention performance, and (c) potential effects of the individual pre-exposure duration on children's referent selection and retention trials (see Appendix D).

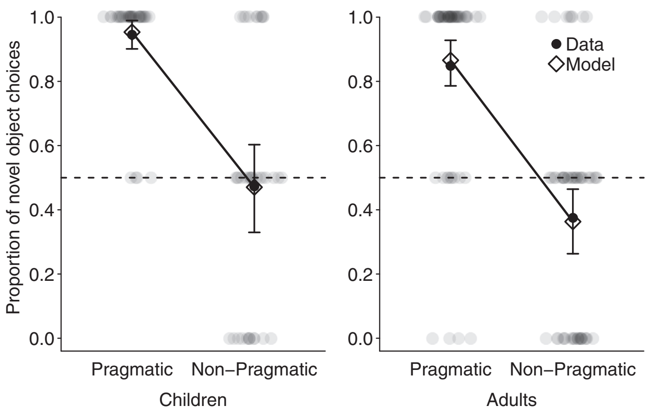

FIGURE 3 Correct choices in referent selection trials by condition and age group. Transparent dots represent the proportions of novel object choices per individual, based on trials in which any selection was made ($n_{children\_Pragmatic} = 73$, $n_{children\_Non\text{-}Pragmatic} = 74$, $n_{adults\_Pragmatic} = 112$, $n_{adults\_Non\text{-}Pragmatic} = 112$). Filled black dots indicate the aggregated proportions of novel object choices per group and diamond shapes the predicted probabilities for a novel object selection by the binomial mixed effects model predicting participants correct choices by age group, condition and their interaction. Vertical lines indicate the 95% confidence intervals. Predicted values and their confidence intervals have been obtained via bootstrapping with 1000 boots. Chance level = 50%.

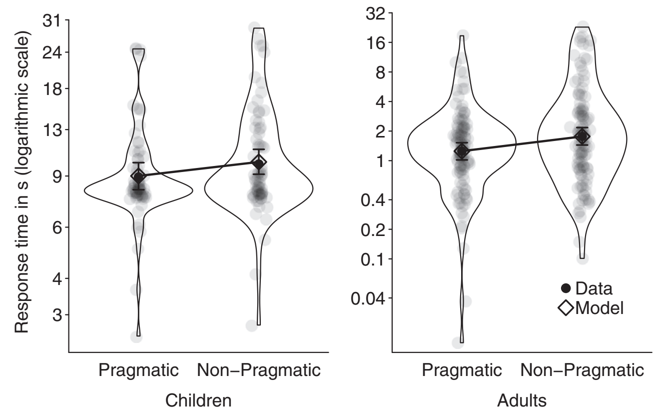

FIGURE 4 Response times by condition and age group. Transparent dots represent response times (in seconds) for each single trial in which any selection was made and in which response times were codable ($n_{children\_Pragmatic} = 72$, $n_{children\_Non\text{-}Pragmatic} = 74$, $n_{adults\_Pragmatic} = 112$, $n_{adults\_Non\text{-}Pragmatic} = 112$). Filled black dots indicate the means of the actual response times and diamond shapes the predicted values by the linear mixed effects models predicting children's and adults' log-transformed reaction times by condition. Vertical lines indicate the 95% confidence intervals. Predicted values and their confidence intervals have been obtained via bootstrapping with 1000 boots. Absolute values of the response times cannot be directly compared for children and adults because response times for children were measured from the start of the speaker's request to show an object and for adults from the end of this request.

3 | RESULTS

3.1 | Object choices in referent selection trials

Children chose the completely novel object significantly above chance in the pragmatic condition ($b = 3.02$, $SE = 0.56$, $p < 0.001$), but their choices did not differ from chance in the non-pragmatic condition ($b = -0.12$, $SE = 0.27$, $p = 0.655$). Adults performed similarly above chance in the pragmatic condition ($b = 1.86$, $SE = 0.33$, $p < 0.001$). However, in the non-pragmatic condition, they selected the novel object *below* chance (instead selecting the pre-exposed object; $b = -0.56$, $SE = 0.23$, $p = 0.014$). The full model including "group" (i.e., each age group and condition combination) as predictor fitted the data significantly better than the null model ($\chi^2 = 92.15$, $df = 3$, $p < 0.001$).

Next, we examined whether performance differed by condition and age group. Both children ($b = -3.14$, $SE = 0.63$, $p < 0.001$) and adults ($b = -2.43$, $SE = 0.43$, $p < 0.001$) were significantly more likely to select the novel object in the pragmatic than the non-pragmatic condition. The age groups did not differ significantly from each other, neither in the pragmatic ($b = -1.16$, $SE = 0.60$, $p = 0.055$) nor in the non-pragmatic condition ($b = -0.44$, $SE = 0.35$, $p = 0.204$). There was no significant interaction between condition and age group ($b = 0.71$, $SE = 0.69$, $p = 0.304$; Figure 3). The full model, including age group, condition and their interaction as predictors, described the data significantly better than the null model ($\chi^2 = 92.15$, $df = 3$, $p < 0.001$).

3.2 | Uncertainty in referent selection trials

First, we investigated differences in participants' logarithmized response times across conditions. For children's log-response times, the model's assumptions of normality and homogeneity were violated. Since simulation studies demonstrated the robustness of mixed effects models in these cases (Schielzeth et al., 2020), revealing potentially imprecise, though unbiased estimates, we kept the model and interpreted its estimates with caution.[10] Children's log-response times did not differ significantly between the non-pragmatic ($M = 10.8$ s, $SD = 5.2$ s) and pragmatic condition ($M = 9.6$ s, $SD = 4.3$ s; $b = 0.11$, $SE = 0.08$, $p = 0.160$). The model on adults' data, however, showed significantly longer log-response times in the non-pragmatic ($M = 3.3$ s, $SD = 4.4$ s) than in the pragmatic condition ($M = 1.9$ s, $SD = 2.4$ s; $b = 0.34$, $SE = 0.15$, $p = 0.022$; Figure 4).

Next, we compared children's social referencing in both conditions. Because the normal distribution of the random effect was violated and the model showed low stability, we interpreted the results with caution. Children showed significantly more frequent social referencing in the non-pragmatic (25.3% of trials) than the pragmatic condition (6.9% of trials; $b = 2.23$, $SE = 0.97$, $p = 0.022$). Thus, our data suggests that both adults and children showed higher uncertainty in the condition that lacked social-pragmatic context, though the effect in both age groups is apparent in different behavioral measures and seems to be more fragile in children.

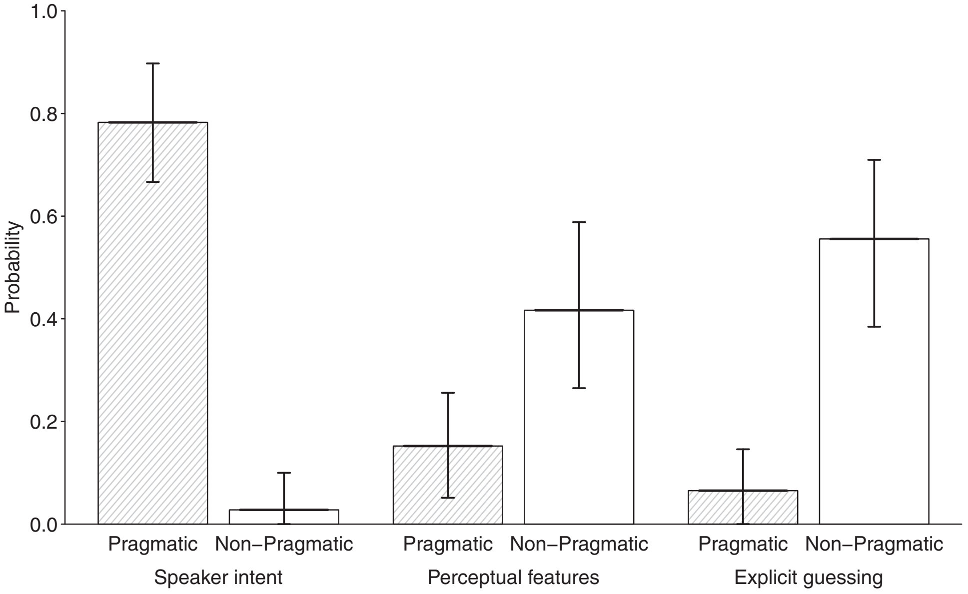

FIGURE 5 Adults' selection strategies across conditions. The bars represent the proportions of actual selection strategies for those adults who indicated strategies matching any of the three depicted categories ($n = 82$; 25 of the 112 participants indicated answers that did not match one of our categories, 20 of which just described their selected object without providing an explanation). The category "experimental logic" was excluded due to its low number (5) of overall observations and its occurrence in only one of both conditions. All 14 of the categorized answers were in line with both the "perceptual features" and the "explicit guessing" category and were assigned to the strategy that got more emphasis (7) or, if they indicated one strategy per trial, to the one related to the first trial (7). Horizontal lines indicate the predicted probability of this strategy by the multinomial model and vertical lines their 95% confidence intervals. Predicted values and their confidence intervals have been obtained via bootstrapping with 1000 boots.

3.3 | Adults' selection strategies (exploratory)

In an exploratory manner, we compared adults' reasoning strategies across conditions. The plot based on our model results (Figure 5) shows that adults in the pragmatic condition were significantly more likely to infer a label's referent based on the speaker's intent than those in the non-pragmatic condition. In contrast, participants in the non-pragmatic condition were significantly more likely than those in the pragmatic condition to indicate their selection was based on guessing or based on perceptual features. Reasoning based on the experimental logic only occurred in the non-pragmatic, not in the pragmatic condition.

3.4 | Object choices in retention trials

Children remembered their previous selections above chance after learning in a pragmatic context (pragmatic condition; $b = 0.58$, $SE = 0.59$, $p = 0.002$), but not when this information was lacking (non-pragmatic condition; $b = -0.32$, $SE = 0.55$, $p = 0.079$). Adults succeeded in retention trials regardless of the learning context: They made consistent choices above chance in the pragmatic ($b = 2.20$, $SE = 0.65$, $p < 0.001$) and the non-pragmatic condition ($b = 2.58$, $SE = 0.69$, $p < 0.001$). The full model including "group" (i.e., each age group and condition combination) as predictor showed significantly better fit than the null model ($\chi^2 = 21.31$, $df = 3$, $p < 0.001$).

Adults showed significantly better retention than children, both in the pragmatic ($b = 1.62$, $SE = 0.81$, $p = 0.047$) and the non-pragmatic condition ($b = 2.90$, $SE = 0.92$, $p = 0.002$; Figure 6). Note, however, that within the group of children, age (treated continuously) did not affect their performance (see exploratory analysis; Appendix D). Condition did not predict participants' retention performance: Consistent choices on retention trials did not differ significantly between conditions, neither for children ($b = -0.90$, $SE = 0.82$, $p = 0.269$) nor for adults ($b = 0.38$, $SE = 0.66$, $p = 0.561$). But note that for this analysis, we cannot rule out statistical power issues, since the study's sample size was determined based on participants' expected disambiguation, not their more fragile retention performance. There was no significant interaction between conditions and age groups ($b = 1.28$, $SE = 1.06$, $p = 0.225$). The model including condition, age group and their interaction as predictors showed a significantly better fit than the null model ($\chi^2 = 21.31$, $df = 3$, $p < 0.001$).

4 | DISCUSSION

The present study aimed to specify the role of object novelty and pragmatic inferences in 2- to 3-year-olds' and adults' referent selection and retention of newly learned words. Based on object choices as well as subtler behavioral measures we examined whether a pragmatic context aids the disambiguation process beyond perceptual cues like object novelty and which circumstances can lead to successful learning

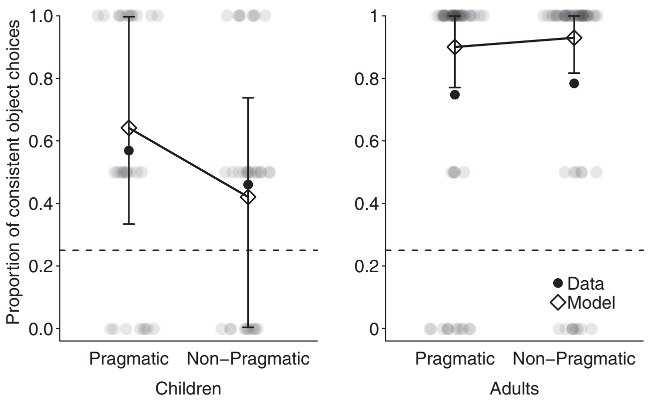

FIGURE 6 Consistent object choices in retention trials by condition and age group. Transparent dots represent the proportions of consistent object choices per individual, based on trials in which any selection was made ($n_{children\_Pragmatic} = 58$, $n_{children\_Non\text{-}Pragmatic} = 62$, $n_{adults\_Pragmatic} = 111$, $n_{adults\_Non\text{-}Pragmatic} = 111$). Filled black dots indicate the aggregated proportions of consistent object choices per group and diamond shapes the predicted probabilities for a consistent object selection by the binomial mixed effects model predicting participants consistent choices by age group, condition and their interaction. Vertical lines indicate the 95% confidence intervals. Predicted values and their confidence intervals have been obtained via bootstrapping with 1000 boots. Chance level = 25%.

of new word-object mappings. Our study revealed three main findings: (1) Young children and adults were able to use pragmatic cues to successfully disambiguate novel words; (2) Both age groups showed more certainty in their selection if presented with a pragmatic context in the word-learning situation; and (3) after 5 min, adults could remember their previous selections regardless of the initial learning situation while children only reliably did so in the pragmatic condition.

4.1 | The role of lexical knowledge in disambiguation

In our design, participants' environments provided only name-unknown objects as potential referents for novel words. We found that even in these situations, that is, without lexical cues present, they were able to find the correct referent by relying on other cues (e.g., common ground information). The results highlight that the nameability of one of the potential referents is not essential for the disambiguation effect. This is in line with past research demonstrating children's ability to disambiguate in the absence of lexical cues (Akhtar et al., 1996; Bohn et al., 2022; Horst et al., 2011; Mather & Plunkett, 2012) and demonstrates that lexical biases alone cannot explain the disambiguation effect completely.

Importantly, this does not imply that lexical knowledge has no beneficial role in the disambiguation process: Both children's overall vocabulary and their specific word knowledge of to-be-excluded referents in the scene have previously been shown to aid their success in resolving lexical ambiguity (Bion et al., 2013; Grassmann et al., 2015; Lewis et al., 2020). Thus, while lexical knowledge is not necessary to solve the disambiguation task, it might nevertheless serve as a beneficial additional cue. It seems plausible to suggest that different processes, instead of demonstrating mutually exclusive options, may actually combine in order to solve referential ambiguity (see e.g., Bohn, Schulze, et al., 2022; Bohn, Tessler, et al., 2021; Frank & Goodman, 2012 for the rational speech act framework and its empirical tests).

4.2 | The role of object novelty in disambiguation

We found that upon hearing a novel word, participants selected the novel (vs. pre-exposed) object *only* if the pragmatic context indicated this to be sensible: When no pragmatic context was given, children selected objects randomly and adults even tended to select the pre-exposed instead of the novel object. Thus, we did not find evidence for children's (or adults') guidance by object novelty alone in situations of referential ambiguity. This result is in line with research highlighting the role of pragmatics in interpreting cues of novelty (Akhtar et al., 1996; Graham et al., 2005; Grassmann et al., 2009; Marno, 2021), but contrasts with studies focusing on the role of internal novelty biases in children's disambiguation (Dysart et al., 2016; Horst et al., 2011; Mather & Plunkett, 2012). In contrast to the latter line of research, the pre-exposure phase in our non-pragmatic condition did not involve another person or human voice that introduced this object to the participants. Thus, here the effect of relative object novelty was not confounded by potential pragmatic effects. This could be one potential reason for the different response pattern of children in the current study.

Other potential explanations concern the overall robustness of the effect (see Mather & Plunkett, 2012, study 2), as well as young children's preference to interact with novel objects regardless of novel words which may explain their selection in studies without a no-label baseline (Dysart et al., 2016; Horst et al., 2011). Further, there may be differences in the relative object salience across studies, which may also affect what grabs children's attention in these situations. There are, for instance, different ways of pre-exposing one object to achieve differing relative object novelty: by jointly playing with it (Dysart et al., 2016; Horst et al., 2011), by a speaker commenting on it (Mather & Plunkett, 2012) or by moving it over the screen (the present study). Recent computational and empirical work has highlighted that children's curiosity and interest may be guided by a complex interplay of factors. In addition to the novelty of a stimulus, these factors also include its salience, complexity and subjective usefulness as well as the learner's prior knowledge and the structure of the learning environment (Dubey & Griffiths, 2020; Kidd & Hayden, 2015; Samuelson et al., 2011). Together, these factors may affect which object children attend to and select, when no information regarding the speakers' referential intent is available.

Future research is needed to investigate under which circumstances the novelty of a stimulus (alone or in combination with further cues, such as object salience or linguistic experience, e.g., Lewis et al., 2020) may drive children's referent selection and how this may change across children's development (Hollich et al., 2000; Yurovsky & Frank, 2017).

An open question is also to what extent children's attraction to novelty is affected by the modality the information is presented in (see Strouse and Samson (2021) on the video deficit phenomenon, but see Frank et al. (2016), Rhodes et al. (2020), and Schidelko et al. (2021) for comparable findings across formats). And finally, while we tried to ensure that the pre-exposure in both conditions was equally engaging and the objects were made equally salient, this is of course always open to debate (see, e.g., Akhtar et al., 1996; Diesendruck et al., 2004; Samuelson & Smith, 1998). Thus, it would be an interesting question for future work to monitor children's focus of attention directly, especially during the pre-exposure phase, by tracking their eye-gaze.

4.3 | The role of pragmatics in disambiguation

Our study revealed that if participants were presented with a novel word, adults and children alike, only selected a novel rather than a pre-exposed object as the correct referent, if the common ground between speaker and listener gave them reason to do so. The comparison of the pragmatic and non-pragmatic conditions suggests that participants' robust disambiguation in the pragmatic context cannot be reduced to their selection being based on relative object novelty. The relative novelty of both potential referents, as well as their contextual salience, were closely matched in both conditions: In both conditions, we ensured that children attended to the pre-exposed object either by movement and sound (non-pragmatic) or by a speaker attending to and commenting on it (pragmatic). Furthermore, slight variance in individual exposure length of the pre-exposed object—brought about by the online presentation format—had no effect on children's performance, neither regarding their initial disambiguation nor their word retention. This pattern of findings suggests that it was indeed the pragmatic cues that supported children's referent disambiguation, and that an attraction to object novelty, by itself, was not sufficient. Moreover, participants' differential behavior in both conditions was not only apparent in their object selections, but also affected their certainty during referent disambiguation. Children showed more frequent social referencing if the labeling event did not provide a pragmatic context to resolve the referential ambiguity. Similarly, adults showed significantly longer reaction times when they could not rely on pragmatic information, indicating their enhanced uncertainty in this context.

Adults' open explanations further supported the pattern found in their response times. While they gave confident explanations in the pragmatic condition based on the speaker's intentions and behavior, they frequently indicated their own ignorance and tried hard to find reasons for selecting one object over the other in the non-pragmatic condition, indicating a high level of monitoring their own uncertainty. Overall, 2- and 3-year-olds and adults showed remarkably similar response patterns in their disambiguation. From an early age on, the common ground that listeners share with a speaker seems to enable them to interpret an otherwise ambiguous lexical environment, enhance their certainty and guide their referent selection.

This set of findings suggests that children as young as two can use pragmatic context to disambiguate novel words: They understand that a speaker's excitement refers to a novel object, rather than a pre-exposed one, if the speaker already attended to the latter. This is in line with previous research highlighting early pragmatic abilities in referent disambiguation (Akhtar et al., 1996; Bohn, Le, et al., 2021; Bohn, Tessler, et al., 2022; Liebal et al., 2009; Marno, 2021; Moll et al., 2006, 2007; Tomasello & Haberl, 2003) and demonstrates the important role of pragmatic processes and common ground information in resolving referential ambiguity. Importantly, we find the same pattern of results in our pre-registered online study using a different scenario, response measure and testing format. This demonstrates a considerable generalizability of the results beyond the details of a certain study. Convergent validity across study designs is of special value in light of the current replication crisis. Future research should address whether these patterns generalize not only across methodologies, but also across cultures.

4.4 | Word learning after successful disambiguation

Both 2- to 3-year-olds and adults were, in principle, able to retain word-object links during a 5-min delay. However, while adults' performance was independent of the initial presentation context (and overall, expectedly higher than children's), children only showed reliable learning in the pragmatic condition. Further, in our exploratory analysis there was no difference in the retention performance of children who selected the novel versus the pre-exposed object in referent selection trials. Thus, while there was no clear evidence that children learned word-object mappings based on relative object novelty (see Kucker et al., 2018, 2020 for similar findings), their referent selection based on pragmatic inferences did indeed demonstrate word learning.

Children's general potential to retain words some minutes after disambiguation (as demonstrated in the pragmatic condition) contributes to a puzzling picture of mixed findings and approaches. Studies have shown both children's success (Carey & Bartlett, 1978; Kalashnikova et al., 2018; Markson & Bloom, 1997; Spiegel & Halberda, 2011; Zosh et al., 2013) as well as their failure (Bion et al., 2013; Horst & Samuelson, 2008; Kucker et al., 2018, 2020; McMurray et al., 2012) in retaining words after successful disambiguation. What might explain this pattern?

Previous debates have stressed (a) that if children are taught just one novel word, effects of attentional highlighting cannot be distinguished from actual word learning and (b) that to assess children's word-learning, retention should not be assessed immediately after referent selection (see Axelsson & Horst, 2013; Horst & Samuelson, 2008). These factors, however, cannot account for children's successful performance in the present study, as we taught children more than one novel word, and implemented a sufficient time delay between the disambiguation and retention trials to test children's long-term memory instead of their repetition of a very recent choice.

Other potential reasons for discrepancies may be age effects and methodological differences. Even though our exploratory analysis revealed no age-related differences in children's retention performance (see Appendix D), the age group tested here (2- to 3-year-olds) was slightly older than those participants who failed to demonstrate word learning in some previous studies (e.g., Horst & Samuelson, 2008; Kucker et al., 2020). Furthermore, it is likely that methodological differences may also have played a role, in particular the number of trials, objects and novel labels children are confronted with during the test session. Thus, our results indicate that, in a context that (a) does not overwhelm children's limited memory capacity and (b) provides at least a minimum of pragmatic discourse information young children, from at least 2 years on, are able to remember word-object mappings after successful disambiguation. Under these pre-conditions, children's referent selection can be demonstrated to be the basis of word learning.

Children's retention performance in the pragmatic condition gives further insights into the role of pragmatics in the word learning process. It demonstrates that common ground information does not only affect children's attentional focus during the conversation (as demonstrated in previous studies; Moll et al., 2006, 2007; Moll & Tomasello, 2007; Tomasello & Haberl, 2003). It provides them with the optimal conditions to both specifically link a certain word to a certain referent and eventually store these words in the long run. This result goes beyond past research and highlights that children's referent selection based on pragmatic discourse information can be the basis for word learning.

4.5 | Conclusion

The current study investigated how object novelty and pragmatic inferences guide referent disambiguation and retention in 2- and 3-year-olds and adults. Upon hearing a novel word, children and adults selected a novel instead of a pre-exposed object as the correct referent if the common ground information between speaker and participant suggested this to be sensible. Importantly, this effect could not be reduced to the attraction by object novelty alone. After 5 min, adults remembered their previous selections above chance. Children, however, only showed reliable learning in the pragmatic condition.

Our results highlight that pragmatics may play an important role in resolving referential ambiguity for proficient language users as well as young word learners. Past research demonstrated that from early on, children have a basic understanding of others as intentional agents (Tomasello, 1999, 2010). We add to this by indicating that children are able to use this understanding to solve referential ambiguity in word learning and enhance their certainty during this process. Our findings show that lexical knowledge and biases are not necessary for the success in disambiguation, while object novelty alone might also not be sufficient to explain children's referent selection. Instead, children can efficiently make sense of referentially ambiguous expressions by relying on their early understanding of others' intentional actions: here, their understanding of others' reactions to new things. The results highlight how our early socio-cognitive capacities do not only help us to infer the meanings of novel words and communicate in the moment. Instead, they present a basis for word learning.

ACKNOWLEDGMENT

This research project has been approved by the ethics committee of the Institute for Psychology, University of Göttingen (project numbers 300 & 301). The research was funded by the Deutsche Forschungsgemeinschaft DFG (254142454/GRK 2070). We want to thank Roger Mundry for statistical support, Marlene Meyer, Lia Künnemann, and Sofya Nechukhaeva for their help with stimuli creation, Mellory Kripzak for help with video coding, and Marlen Kaufmann for help with organizing the data collection.

Open access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST

The authors have no conflict of interest.

DATA AVAILABILITY STATEMENT

The data set, descriptive data, R scripts, detailed model results and assumption tests are accessible on OSF (https://osf.io/bk47x/). This study was preregistered on OSF; see https://osf.io/p2re6.

ENDNOTES

[i] For their second experiment, Mather and Plunkett (2012) conducted separate analyses for each of the three trials infants participated in and an overall analysis based on the averages across trials. They found that infants' attention to the completely novel object did not increase after the onset of naming on the first or second trial, but did so on their third trial. The overall analysis confirmed that in experiment 2 infants' attention to the completely novel object did not increase after the onset of naming.

[ii] On average, this led to slightly longer (though overlapping) presentation times of the pre-exposed object in the pragmatic than the non-pragmatic condition: The average actual presentation time of the pre-exposed object (before the novel object appeared) was 69.3 s (SD = 38.8 s) for the Pragmatic and 33.8 (SD = 9.5 s) for the non-pragmatic condition. The novel object was presented for M = 28.6 s (SD = 7.4 s) in the pragmatic and M = 29.7 s (SD = 9.0 s) in the non-pragmatic condition. Importantly, however, an additional analysis confirmed that the exact presentation time of the pre-exposed object had no effect on children's performance, neither in the referent selection nor the retention trials.

[iii] By playing background music, we ensured that (a) children would continue looking at the screen during the otherwise little engaging pre-exposure and (b) there is a non-social equivalent for the attention brought toward the pre-exposed object by the speaker in the pragmatic condition.

[iv] Webcam recordings of 32 referent selection trials (>20%) were coded. One video was not codable due to missing sound: The child pointed at different locations but was talking to the parent during the selection (not audible to the coder), possibly explaining the selection.

[v] We preregistered to measure the time between the end of the speaker's request to show an object until the child's reaction but realized children's tendency to respond before the speaker finished her request (which was only after ca. 10 s and after the third mention of the target label). To discriminate between those otherwise 0-response times, we started the measurement at label onset. Note that none of the children responded before the label was mentioned for the first time and, thus, that all responses indicated direct reactions to the label.

[vi] We used the following key packages and functions for analyses: glmer() from the package lme4 version 1.1-27.1 (Bates et al., 2015) for binomial mixed effects models, glht() from the package multcomp version 1.4-17 (Hothorn et al., 2008) to test the binomial outcome against a chance level other than 0.5, and lmer() from the package lmerTest version 3.1-3 (Kuznetsova et al., 2017) for linear mixed effects models. See Appendix C for the full list of packages and functions used.

[vii] By suppressing the intercept, the model calculates no estimate for the intercept, but four estimates for the four groups, representing the logit of the probability of the referring group to select the novel object. It then tests each of them against a value of 0 in the model's logit-space, which corresponds to a probability of 0.5 (logit(0.5) = 0).

[viii] We preregistered to test the main effects for both options of reference groups in an additional step of contrast analyses. However, since the same information can be obtained from the same model by recoding the factors' reference groups, we skipped this step to have a simpler and more accurate analysis. We applied the same logic to the corresponding retention model.

[ix] We preregistered to analyze children's and adults' response times combined in one model. However, after changing the coded time frame for children (starting from label onset), response times for both age groups were not comparable anymore. Therefore, we decided to analyze them in separate models.

[x] Note that we additionally ran a model based on a gamma distribution (i.e., assuming no homogenous or normally distributed residuals) and obtained similar results.

REFERENCES

Akhtar, N., Carpenter, M., & Tomasello, M. (1996). The role of discourse novelty in early word learning. *Child Development*, *67*(2), 635–645. https://doi.org/10.2307/1131837

Au, T. K., & Glusman, M. (1990). The principle of mutual exclusivity in word learning: To honor or not to honor? *Child Development*, *61*(5), 1474–1490. https://doi.org/10.2307/1130757

Axelsson, E. L., & Horst, J. S. (2013). Testing a word is not a test of word learning. *Acta Psychologica*, *144*(2), 264–268. https://doi.org/10.1016/j.actpsy.2013.07.002

Bache, S. M., & Wickham, H. (2020). magrittr: A forward-pipe operator for R. https://CRAN.R-project.org/package=magrittr

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bion, R. A. H., Borovsky, A., & Fernald, A. (2013). Fast mapping, slow learning: Disambiguation of novel word–object mappings in relation to vocabulary learning at 18, 24, and 30 months. *Cognition*, *126*(1), 39–53. https://doi.org/10.1016/j.cognition.2012.08.008

Bohn, M., & Frank, M. C. (2019). The pervasive role of pragmatics in early language. *Annual Review of Developmental Psychology*, *1*(1), 223–249. https://doi.org/10.1146/annurev-devpsych-121318-085037

Bohn, M., Le, K. N., Peloquin, B., Köymen, B., & Frank, M. C. (2021). Children's interpretation of ambiguous pronouns based on prior discourse. *Developmental Science*, *24*(3), e13049. https://doi.org/10.1111/desc.13049

Bohn, M., Schulze, C., Schmidt, L. S., Frank, M. C., & Tessler, M. H. (2022). Modeling individual differences in children's information integration during pragmatic word learning. *Open Mind*, *6*, 311–326. https://doi.org/10.1162/opmi_a_00069

Bohn, M., Tessler, M. H., Merrick, M., & Frank, M. C. (2021). How young children integrate information sources to infer the meaning of words. *Nature Human Behaviour*, *5*(8), 1046–1054. https://doi.org/10.1038/s41562-021-01145-1

Bohn, M., Tessler, M. H., Merrick, M., & Frank, M. C. (2022). Predicting pragmatic cue integration in adults' and children's inferences about novel word meanings. *Journal of Experimental Psychology: General*, *151*(11), 2927–2942. https://doi.org/10.1037/xge0001216

Carey, S., & Bartlett, E. (1978). Acquiring a single new word. *Papers and Reports on Child Language Development*, *15*, 17–29.

Clark, E. V. (2015). Common ground. In B. MacWhinney & W. O'Grady (Eds.), *The handbook of language emergence* (pp. 328–353). John Wiley & Sons. https://doi.org/10.1002/9781118346136.ch15

Diesendruck, G., & Markson, L. (2001). Children's avoidance of lexical overlap: A pragmatic account. *Developmental Psychology*, *37*(5), 630–641. https://doi.org/10.1037/0012-1649.37.5.630

Diesendruck, G., Markson, L., Akhtar, N., & Reudor, A. (2004). Two-year-olds' sensitivity to speakers' intent: An alternative account of Samuelson and Smith. *Developmental Science*, *7*(1), 33–41. https://doi.org/10.1111/j.1467-7687.2004.00320.x

Dubey, R., & Griffiths, T. L. (2020). Reconciling novelty and complexity through a rational analysis of curiosity. *Psychological Review*, *127*(3), 455–476. https://doi.org/10.1037/rev0000175

Dysart, E. L., Mather, E., & Riggs, K. J. (2016). Young children's referent selection is guided by novelty for both words and actions. *Journal of Experimental Child Psychology*, *146*, 231–237. https://doi.org/10.1016/j.jecp.2016.01.003

Forstmeier, W., & Schielzeth, H. (2011). Cryptic multiple hypotheses testing in linear models: Overestimated effect sizes and the winner's curse. *Behavioral Ecology and Sociobiology*, *65*(1), 47–55. https://doi.org/10.1007/s00265-010-1038-5

Fox, J., & Weisberg, S. (2019). *An R companion to applied regression (Third)*. Sage. https://socialsciences.mcmaster.ca/jfox/Books/Companion/

Frank, M. C., Bergelson, E., Bergmann, C., Cristia, A., Floccia, C., Gervain, J., Hamlin, J. K., Hannon, E. E., Kline, M., Levelt, C., Lew-Williams, C., Nazzi, T., Panneton, R., Rabagliati, H., Soderstrom, M., Sullivan, J., Waxman, S., & Yurovsky, D. (2017). A collaborative approach to infant research: Promoting reproducibility, best practices, and theory-building. *Infancy*, *22*(4), 421–435. https://doi.org/10.1111/infa.12182

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

Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, *336*(6084), 998–998. https://doi.org/10.1126/science.1218633

Frank, M. C., Sugarman, E., Horowitz, A. C., Lewis, M. L., & Yurovsky, D. (2016). Using tablets to collect data from young children. *Journal of Cognition and Development*, *17*(1), 1–17. https://doi.org/10.1080/15248372.2015.1061528

Goeke, C., Finger, H., Diekamp, D., Standvoss, K., & König, P. (2017). LabVanced: A unified javascript framework for online studies. In *International conference on computational social science (Cologne)*.

Golinkoff, R. M., Hirsh-Pasek, K., Bailey, L. M., & Wenger, N. R. (1992). Children and adults use lexical principles to learn new nouns. *Developmental Psychology*, *28*(1), 99–108. https://doi.org/10.1037/0012-1649.28.1.99

Golinkoff, R. M., Mervis, C., & Hirsh-Pasek, K. (1994). Early object labels: The case for a developmental lexical principles framework. *Journal of Child Language*, *21*(1), 125–155. https://doi.org/10.1017/S0305000900008692

Graham, S. A., Turner, J. N., & Henderson, A. M. E. (2005). The influence of object pre-exposure on two-year-olds' disambiguation of novel labels. *Journal of Child Language*, *32*(1), 207–222. https://doi.org/10.1017/S030500090400666X

Grassmann, S., Schulze, C., & Tomasello, M. (2015). Children's level of word knowledge predicts their exclusion of familiar objects as referents of novel words. *Frontiers in Psychology*, *6*, 1200. https://doi.org/10.3389/fpsyg.2015.01200

Grassmann, S., Stracke, M., & Tomasello, M. (2009). Two-year-olds exclude novel objects as potential referents of novel words based on pragmatics. *Cognition*, *112*(3), 488–493. https://doi.org/10.1016/j.cognition.2009.06.010

Grolemund, G., & Wickham, H. (2011). Dates and times made easy with lubridate. *Journal of Statistical Software*, *40*(3), 1–25.

Halberda, J. (2003). The development of a word-learning strategy. *Cognition*, *87*(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*(4), 310–344. https://doi.org/10.1016/j.cogpsych.2006.04.003

Hollich, G. J., Hirsh-Pasek, K., Golinkoff, R. M., Brand, R. J., Brown, E., Chung, H. L., Hennon, E., Rocroi, C., & Bloom, L. (2000). Breaking the language barrier: An emergentist coalition model for the origins of word learning. *Monographs of the Society for Research in Child Development*, *65*(3), i–135.

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., Samuelson, L. K., Kucker, S. C., & McMurray, B. (2011). What's new? Children prefer novelty in referent selection. *Cognition*, *118*(2), 234–244. https://doi.org/10.1016/j.cognition.2010.10.015

Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal*, *50*(3), 346–363.

Kalashnikova, M., Escudero, P., & Kidd, E. (2018). The development of fast-mapping and novel word retention strategies in monolingual and bilingual infants. *Developmental Science*, *21*(6), e12674. https://doi.org/10.1111/desc.12674

Kalashnikova, M., Mattock, K., & Monaghan, P. (2016). Mutual exclusivity develops as a consequence of abstract rather than particular vocabulary knowledge. *First Language*, *36*(5), 451–464. https://doi.org/10.1177/0142723716648850

Kidd, C., & Hayden, B. Y. (2015). The psychology and neuroscience of curiosity. *Neuron*, *88*(3), 449–460. https://doi.org/10.1016/j.neuron.2015.09.010

Kucker, S. C., McMurray, B., & Samuelson, L. K. (2018). Too much of a good thing: How novelty biases and vocabulary influence known and novel referent selection in 18-month-old children and associative learning models. *Cognitive Science*, *42*(S2), 463–493. https://doi.org/10.1111/cogs.12610

Kucker, S. C., McMurray, B., & Samuelson, L. K. (2020). Sometimes it is better to know less: How known words influence referent selection and retention in 18- to 24-month-old children. *Journal of Experimental Child Psychology*, *189*, 104705. https://doi.org/10.1016/j.jecp.2019.104705

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Lewis, M., Cristiano, V., Lake, B. M., Kwan, T., & Frank, M. C. (2020). The role of developmental change and linguistic experience in the mutual exclusivity effect. *Cognition*, *198*, 104191. https://doi.org/10.1016/j.cognition.2020.104191

Liebal, K., Behne, T., Carpenter, M., & Tomasello, M. (2009). Infants use shared experience to interpret pointing gestures. *Developmental Science*, *12*(2), 264–271. https://doi.org/10.1111/j.1467-7687.2008.00758.x

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

Markson, L., & Bloom, P. (1997). Evidence against a dedicated system for word learning in children. *Nature*, *385*, 813–815. https://doi.org/10.1038/385813a0

Marno, H. (2021). Is 2-year-old children's referential disambiguation of a novel word primarily aimed at identifying the word meaning or at understanding their communicative partner? *Developmental Psychology*, *57*(7), 1058–1066. https://doi.org/10.1037/dev0001204

Mather, E., & Plunkett, K. (2012). The role of novelty in early word learning. *Cognitive Science*, *36*(7), 1157–1177. https://doi.org/10.1111/j.1551-6709.2012.01239.x

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

Mervis, C. B., & Bertrand, J. (1994). Acquisition of the novel name-nameless category (N3C) principle. *Child Development*, *65*(6), 1646–1662. https://doi.org/10.2307/1131285

Moll, H., Carpenter, M., & Tomasello, M. (2007). Fourteen-month-olds know what others experience only in joint engagement. *Developmental Science*, *10*(6), 826–835. https://doi.org/10.1111/j.1467-7687.2007.00615.x

Moll, H., Koring, C., Carpenter, M., & Tomasello, M. (2006). Infants determine others' focus of attention by pragmatics and exclusion. *Journal of Cognition and Development*, *7*(3), 411–430. https://doi.org/10.1207/s15327647jcd0703_9

Moll, H., & Tomasello, M. (2007). How 14- and 18-month-olds know what others have experienced. *Developmental Psychology*, *43*(2), 309–317. https://doi.org/10.1037/0012-1649.43.2.309

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*, 195–203. https://doi.org/10.3758/s13428-018-01193-y

Phillips, N. (2017). Yarrr: A companion to the e-book "YaRrr!: The pirate's guide to R". https://CRAN.R-project.org/package=yarrr

Pomiechowska, B., Bródy, G., Csibra, G., & Gliga, T. (2021). Twelve-month-olds disambiguate new words using mutual-exclusivity inferences. *Cognition*, *213*, 104691. https://doi.org/10.1016/j.cognition.2021.104691

Quine, W. V. O. (1960). *Word and object: An inquiry into the linguistic mechanisms of objective reference* (pp. xv, 294). MIT Press.

R Core Team. (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, https://www.R-project.org/

Rhodes, M., Rizzo, M. T., Foster-Hanson, E., Moty, K., Leshin, R. A., Wang, M., Benitez, J., & Ocampo, J. D. (2020). Advancing developmental science via unmoderated remote research with children. *Journal of Cognition and Development*, *21*(4), 477–493. https://doi.org/10.1080/15248372.2020.1797751

Samuelson, L. K., & Smith, L. B. (1998). Memory and attention make smart word learning: An alternative account of Akhtar, Carpenter, and Tomasello. *Child Development*, *69*(1), 94–104. https://doi.org/10.1111/j.1467-8624.1998.tb06136.x

Samuelson, L. K., Smith, L. B., Perry, L. K., & Spencer, J. P. (2011). Grounding word learning in space. *PLOS ONE*, *6*(12), e28095. https://doi.org/10.1371/journal.pone.0028095

Schidelko, L. P., Schünemann, B., Rakoczy, H., & Proft, M. (2021). Online testing yields the same results as lab testing: A validation study with the false belief task. *Frontiers in Psychology*, *12*, 4573. https://www.frontiersin.org/article/10.3389/fpsyg.2021.703238

Schielzeth, H., Dingemanse, N. J., Nakagawa, S., Westneat, D. F., Allegue, H., Teplitsky, C., Réale, D., Dochtermann, N. A., Garamszegi, L. Z., & Araya-Ajoy, Y. G. (2020). Robustness of linear mixed-effects models to violations of distributional assumptions. *Methods in Ecology and Evolution*, *11*(9), 1141–1152. https://doi.org/10.1111/2041-210X.13434

Simons, D. J., & Rensink, R. A. (2005). Change blindness: Past, present, and future. *Trends in Cognitive Sciences*, *9*(1), 16–20. https://doi.org/10.1016/j.tics.2004.11.006

Smith, L. B. (2000). Learning how to learn words: An associative crane. In R. M. Golinkoff, K. Hirsh-Pasek, L. Bloom, L. B. Smith, A. Woodward, N. Akhtar, M. Tomasello, & G. Hollich (Eds.), *Becoming a word learner: A debate on lexical acquisition* (pp. 51–80). Oxford University Press.

Spiegel, C., & Halberda, J. (2011). Rapid fast-mapping abilities in 2-year-olds. *Journal of Experimental Child Psychology*, *109*(1), 132–140. https://doi.org/10.1016/j.jecp.2010.10.013

Strouse, G. A., & Samson, J. E. (2021). Learning from video: A meta-analysis of the video deficit in children ages 0 to 6 years. *Child Development*, *92*(1), e20–e38. https://doi.org/10.1111/cdev.13429

Tomasello, M. (1999). Having intentions, understanding intentions, and understanding communicative intentions. In P. D. Zelazo, J. W. Astington, & D. R. Olson (Eds.), *Developing theories of intention: Social understanding and self-control* (pp. 63–75). Lawrence Erlbaum Associates Publishers.

Tomasello, M. (2010). *Origins of human communication*. MIT Press.

Tomasello, M., & Haberl, K. (2003). Understanding attention: 12- and 18-month-olds know what is new for other persons. *Developmental*

Psychology, *39*(5), 906–912. https://doi.org/10.1037/0012-1649.39.5.906

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S* (Fourth). Springer. https://www.stats.ox.ac.uk/pub/MASS4/

Warnes, G. R., Bolker, B., & Lumley, T. (2021). gtools: Various R programming tools. https://CRAN.R-project.org/package=gtools

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., & Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Yurovsky, D., & Frank, M. C. (2017). Beyond naïve cue combination: Salience and social cues in early word learning. *Developmental Science*, *20*(2), e12349. https://doi.org/10.1111/desc.12349

Zosh, J. M., Brinster, M., & Halberda, J. (2013). Optimal contrast: Competition between two referents improves word learning. *Applied Developmental Science*, *17*(1), 20–28. https://doi.org/10.1080/10888691.2013.748420

How to cite this article: Bleijlevens, N., Contier, F., & Behne, T. (2023). Pragmatics aid referent disambiguation and word learning in young children and adults. *Developmental Science*, e13363. https://doi.org/10.1111/desc.13363

APPENDIX A

PARTICIPANT EXCLUSION

In addition to the final sample, 34 children participated, but were excluded from analyses based on our pre-registered exclusion criteria: at least one incorrect answer in practice-II trials (14), multilingual language acquisition (6), uncooperative behavior (3), and technical issues at an early point of the experiment (11). Further, six additional children provided data for referent selection, but not retention trials, due to technical issues toward the end of the session, and these were included for the analyses of referent selection trials.

As an additional pre-registered exclusion criteria, we coded children's substantial distraction and parents' interference in the selection process for referent selection and retention trials in a binary manner. A trained, blinded coder pre-screened children's overall attention via their webcam videos and in case of partial distraction, it was decided whether the child was distracted during crucial scenes (i.e., during an object's appearance/movement or the speaker's appearance/disappearance/speech). Reliability coding by a second blinded coder for 32 referent selection trials (>20%) revealed 100% agreement for distraction and 97% agreement for parental interference coding. In the single case of disagreement for parental interference, the parent waited for a second pointing action; however, as the first pointing was probably directed at the bear instead of any object (based on the parent saying "Yes. Bear"), we finally did not count this as a case of interference.

Based on this, we excluded nine single trials within the final sample of children: For one child, one referent selection trial was excluded due to distraction during crucial scenes, and as a consequence, the corresponding retention trial was excluded too; for two children, both retention trials were excluded because of parental interference in the first one; and for three children, only their second retention trial was excluded due to parental interference in this trial.

APPENDIX B

COUNTERBALANCING AND RANDOMIZATION

In the practice-I phase, the object pairings presented (apple + teddy, car + bunny, flower + cat, bus + doll), the target objects (apple, car, flower, bus), trial order (apple—car—flower—bus) and target locations (left—right—right—left) were fixed.

In the practice-II phase, the object pairings presented (ball + duck, shoe + banana) and the target objects (ball and shoe, respectively) were pre-defined, but we randomized which of these object pairings was presented on the *first* trial and on which side the target was located. For the second trial, the platform chose the remaining object pair and the target was located on the opposite side. Within participants, target locations were fully counterbalanced.

In the referent selection phase, we randomized for the *first* referent selection trial: (a) which of the two pre-defined novel object-pairs (the "modi"-pair vs. "toma"-pair) was presented (which went along with asking for this label), (b) the target location (left vs. right), and (c) only for the pragmatic condition: which table bear turned to first (the one with the pre-exposed object vs. the empty one). For the second trial, the platform chose the opposite option for each variable, thereby counterbalancing each of these factors for each participant. Additionally, we fully counterbalanced within each age group (children and adults) which object within a given object pair was the target.

For the retention phase, we created eight counterbalanced options for object locations in both trials. Across these options, each object was presented equally often at each position. For each participant, each object changed its position from the first to the second trial. We randomized (a) which option was chosen for a certain participant and (b) which label was asked for first (modi first/toma first).

APPENDIX C

PACKAGES AND FUNCTIONS

We used the following packages for data handling, preparation, and visualization: tidyverse version 1.3.1 (Wickham et al., 2019), magrittr version 2.0.1 (Bache & Wickham, 2020), lubridate version 1.8.0 (Grolemund & Wickham, 2011), and yarrr version 0.1.5 (Phillips, 2017).

For data analysis, we used the following packages and functions: glmer() from the package lme4 version 1.1-27.1 (Bates et al., 2015) for binomial mixed effects models, glht() from the package multcomp version 1.4-17 (Hothorn et al., 2008) to test the binomial outcome against a chance level other than 0.5, lmer() from the package lmerTest version 3.1-3 (Kuznetsova et al., 2017) for linear mixed effects models, multinom() from the package nnet version 7.3-16 (Venables & Ripley, 2002) for the (exploratory) multinomial model, mvrnorm() from the package MASS version 7.3-54 (Venables & Ripley, 2002) and logit() from the package gtools version 3.9.2 (Warnes et al., 2021) for parts of the data simulation, and vif() from the package car version 3.0-12 (Fox & Weisberg, 2019) to calculate variance inflation factors.

APPENDIX D

EXPLORATORY ANALYSES

Adults' Selection Strategies

To test for differences in adults' strategies across conditions, we fitted a multinomial model and predicted their strategies by condition. To avoid problems with the model's stability, we excluded low-frequent strategies that were indicated in only one of both conditions from this analysis. Based on this model, we then extracted and plotted the predicted probabilities for the occurrence of each strategy per condition, as well as their 95% confidence intervals. We used this plot for a visual analysis and interpretation of the results. For the results, see main paper and Figure 5.

Children's Retention Performance

We also ran two exploratory analyses to examine children's retention performance. First, to test for age effects in children's performance in retention trials, we ran a binomial mixed effects model with children's age (continuous), condition, and their interaction as predictors and with random intercepts for participants. We log-transformed age due to theoretical considerations (since the same age differences are expected to show stronger developmental changes in younger than in older children) and z-transformed it afterward. The model revealed that children's performance on the retention trials did not change with age, neither in the pragmatic ($b = -0.01$, $SE = 0.29$, $p = 0.976$) nor in the non-pragmatic condition ($b = -0.27$, $SE = 0.27$, $p = 0.301$; see Figure D1).

Second, for the non-pragmatic condition, we tested whether children's object choice in the referent selection trials (novel vs. pre-exposed) affected their subsequent retention performance, allowing us to compare our results with previous studies who only tested retention in participants who passed the referent selection trials (e.g., Horst & Samuelson, 2008). Thus, we ran a binomial mixed effects models on children's consistent choices in non-pragmatic retention trials, including their previous choice in referent selection (novel vs. pre-exposed object) as a predictor and random intercepts for participants. We found that children's retention performance in the non-pragmatic condition did not differ depending on which object they selected in the previous referent selection trial ($b = -0.21$, $SE = 0.51$, $p = 0.682$). We did not include the pragmatic condition in this analysis, because here children selected the pre-exposed object on only four referent selection trials in total. In none of these did they select this object again in the corresponding retention trial.

Effects of the Individual Pre-exposure Duration

Finally, to test for potential effects of the individual pre-exposure duration on referent selection or retention performance, we ran two exploratory binomial mixed effects models, one on children's referent selection and one on their retention performance. We included condition, the actual presentation time of the pre-exposed object on the screen and their interaction as predictors and added random intercepts for participants. We log-transformed the durations for conceptual reasons (since an increase of 1 s is expected to have a larger impact for shorter than for longer durations) and z-transformed them afterward. The models revealed that children's selections were not affected by the length of exposure to the pre-exposed object, neither for referent selection trials (pragmatic condition: $b = 0.94$, $SE = 1.08$, $p = 0.386$; non-pragmatic condition: $b = 0.03$, $SE = 0.59$, $p = 0.955$) nor retention trials (pragmatic condition: $b = 0.25$, $SE = 0.31$, $p = 0.416$; non-pragmatic condition: $b = 4.69$, $SE = 4.27$, $p = 0.272$).

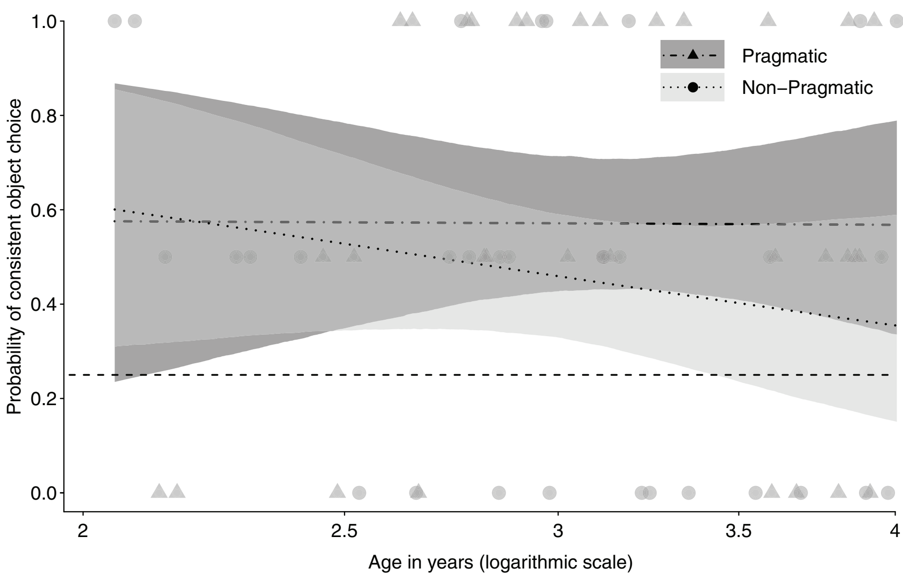

FIGURE D1 Children's consistent choices in retention trials by age and condition. Transparent triangles (pragmatic) and dots (non-pragmatic) represent the proportions of consistent object choices per child, based on trials in which any selection was made ($n_{\text{Pragmatic}} = 58$, $n_{\text{Non-Pragmatic}} = 62$). Dash-dotted (pragmatic) and dotted (non-pragmatic) lines indicate the fitted values obtained by the binomial mixed effects model predicting children's consistent choices by log-transformed age (continuous), condition and their interaction. Darker (pragmatic) and lighter (non-pragmatic) gray polygons show the 95% confidence intervals. Fitted values and their confidence intervals have been obtained via bootstrapping with 10,000 boots. Chance level = 25%.