When language background does not matter:

Bilingual children disambiguate and learn novel words just like their monolingual peers

Natalie Bleijlevens[1,2], Anna-Lena Ciesla[1] and Tanya Behne[1,2]

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

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

Author Note

Natalie Bleijlevens https://orcid.org/0000-0002-4421-4153

Anna-Lena Ciesla

Tanya Behne https://orcid.org/0000-0001-7649-4949

This study was preregistered on OSF; see https://osf.io/x9aej. The complete data, analysis code and experimental stimuli are accessible there as well (https://osf.io/9epuw/). This research project has been approved by the ethics committee of the Institute for Psychology, University of Göttingen (project number 317b). There are no conflicts of interest to disclose. The research was funded by the Deutsche Forschungsgemeinschaft DFG (254142454/GRK 2070). We want to thank Roger Mundry for statistical support. Thanks to Lucy Schmitz and Tina Zahrie for their help with the study implementation, Lorna Schlüter for her help with coding, and Marlene Meyer, Lia Künnemann, and Sofya Nechukhaeva for their help with the stimuli creation.

Correspondence concerning this article should be addressed to Natalie Bleijlevens, University of Göttingen, Waldweg 26, 37073 Göttingen, Germany. Email: natalie.bleijlevens@uni-goettingen.de

Research Highlights

- We tested if mono- and bilingual preschoolers (and adults) differ in the strategies and pragmatic skills used to disambiguate and learn novel words.
- Both language groups successfully disambiguated and retained words in both a mutual exclusivity task and a task that required interpreting the pragmatic context.
- In comparable samples of mono- and bilinguals, their language background does not seem to impact how they learn the meanings of novel words.
- These findings contribute to ongoing theoretical debates about whether (and, if so, why) mono- and bilinguals might differ in their word-learning strategies.

Abstract

Do mono- and bilingual children differ in the way they learn novel words in ambiguous settings? Listeners may resolve referential ambiguity by assuming that novel words refer to unknown, rather than known, objects –a response known as the *mutual exclusivity effect*. Past research suggested a bilingual disadvantage with regard to this disambiguation strategy, perhaps because, across languages, bilinguals' experience contradicts one-to-one mappings of label and referent. Another line of research suggested a bilingual advantage in resolving referential ambiguity, based on bilinguals' advanced pragmatic skills. Here, we examine both these claims in a preregistered study with comparable samples of mono- and bilingual 3-year-olds (n=74) and adults (n=86). We tested referent disambiguation and retention in two tasks: In the Mutual-Exclusivity task, a speaker used a novel label in presence of a known and an unknown object. In the Pragmatic task, she used another novel label in presence of two unknown objects and participants could infer from the pragmatic context that the speaker referred to the object that was new in their discourse. Mono- and bilinguals were equally successful in inferring the correct label-referent links in both tasks and retained them after a delay. These findings indicate that children with different language backgrounds can develop the same strategies and pragmatic skills to learn novel words, highlighting the importance of testing comparable samples of mono- and bilinguals. Children can use their lexical knowledge and socio-cognitive skills to infer the meanings of novel words, irrespective of whether they are acquiring one or more languages.

Keywords: disambiguation, mutual exclusivity, bilingualism, word learning, pragmatics, common ground

When language background does not matter: Bilingual children disambiguate and learn novel words just like their monolingual peers

Young language learners face many challenges. Based on limited vocabulary and language experience, they need to find out what their conversation partner is referring to by using a novel word. Crucially, they often need to do so while facing referential ambiguity, i.e., in the presence of several potential referents. Despite these challenges, children come to learn the words they are exposed to; and they do so regardless of whether they are learning only one or several native languages. Mono- and multilingual children differ fundamentally in their language experience – But do they also differ in the strategies they lean on to learn novel word meanings?

One line of research proposes that bilinguals have *disadvantages* in applying strategies that monolinguals use to disambiguate novel word meanings. For example, listeners may resolve referential ambiguity by assuming that a novel label refers to a novel rather than a name-known object – a strategy known as the "mutual exclusivity" (ME) effect (see e.g., Lewis et al., 2020). While monolingual children reliably use this strategy, research with bilingual children suggests that they are less likely to do so (e.g., Byers-Heinlein & Werker, 2009). A second line of research proposes that bilingual have *advantages* in disambiguating novel word meaning. For example, it has been suggested that the demands of growing up in a bilingual environment result in superior pragmatic abilities that may in turn aid children's referent identification (Siegal et al., 2009; Wermelinger et al., 2017; Yow et al., 2017; Yow & Markman, 2015). We will examine these two lines of research in turn.

Bilingual Disadvantages in Referent Disambiguation

Children often need to find the referents of novel words in the light of referential ambiguity, i.e., when several potential referents are present. Children, at least monolinguals, have been demonstrated to resolve this ambiguity by showing the mutual exclusivity (ME) effect: They assume that the novel label refers to a novel/label-unknown, rather than to a familiar/label-known, object (Lewis et al., 2020; Markman & Wachtel, 1988), thereby preferring to assign only one label to concepts (i.e., one-to-one mappings). But why might children's language background, that is, growing up in a mono- vs. bilingual environment, affect how they respond (e.g., Byers-Heinlein & Werker, 2009)? Why might bilingual children be less likely to show the ME effect? Two different types of explanations have been put forward.

The first is that mono- and bilingual children use fundamentally different strategies to learn novel words. It has been proposed that bilingual children do not develop a ME strategy due to their language experience and/or the structure of their mental lexicons (e.g., Byers-Heinlein & Werker, 2009; Houston-Price et al., 2010). Bilinguals' regular experience with more than one language contradicts the strict regularity of one-to-one-mappings, at least across languages. As their lexicon includes an increasing number of cross-language synonyms or translation equivalents, they may be more willing to accept many-to-one mappings (Byers-Heinlein & Werker, 2009, 2013; Davidson et al., 1997). In line with this, some research showed that the deficit in using ME was specific to those bilinguals who knew many translation equivalents (Byers-Heinlein & Werker, 2013). However, the empirical evidence is sparse, and other studies failed to find the suggested relationship (e.g., Frank & Poulin-Dubois, 2002; Houston-Price et al., 2010; Weatherhead et al., 2021).

The second type of explanation is that mono- and bilingual children have the same strategies available, but bilinguals may not always be able to apply them for various reasons. For example, bilinguals' vocabulary development within each language is often delayed compared to that of their monolingual peers (e.g., Bialystok et al., 2010). If bilinguals' knowledge of the familiar distractor labels is more fragile, this may result in a weaker ME effect (Grassmann et al., 2015; Lewis et al., 2020). In addition, other performance issues may prevent bilinguals from demonstrating their full capacities: e.g., they may be less comfortable with the test language than monolinguals, and thus perform less proficiently.

However, irrespective of *how* potential differences in ME performance between monolingual and bilingual children may be explained, the more fundamental question that needs examining is *whether* the reported difference is a robust effect. Do mono- and bilingual children really differ in their ME performance? A closer look at the empirical findings reveals a mixed pattern: While some studies find bilinguals to show a weaker ME effect than monolinguals (Byers-Heinlein & Werker, 2009; Houston-Price et al., 2010; Repnik et al., 2021), other studies could not replicate this difference (Byers-Heinlein et al., 2014; I. Frank & Poulin-Dubois, 2002; Kalashnikova et al., 2015, 2018; Rocha-Hidalgo et al., 2021; Rochanavibhata et al., 2022; Weatherhead et al., 2021). A meta-analysis, including 12 studies with multilingual participants, suggested that the magnitude of the ME effect is influenced by the participants' language background as well as their age: The effect sizes tended to be larger for monolingual and older children (Lewis et al., 2020). This analysis, however, did not include more recent research (published after 2017) – much of which did not find a difference in ME performance between mono- and bilingual children (e.g., Kalashnikova et al., 2018; Rocha-Hidalgo et al., 2021; Rochanavibhata et al., 2022; Weatherhead et al., 2021).

Some of these recent studies have suggested a new twist on how word learning in ME contexts may differ between mono- and bilinguals. They distinguished between the use of ME to identify the intended object in the moment of interaction and the long-term retention of this label-object link. Thus, in addition to children's immediate referent selection these studies also assessed children's retention of the novel label-referent link. And while in these studies bilinguals showed the ME effect in their referent selections just as their monolingual peers did, bilingual two-year-olds did not succeed in subsequent retention of labels encountered in a ME context (Kalashnikova et al., 2018; Repnik et al., 2021; Rocha-Hidalgo et al., 2021; but see Weatherhead et al., 2021). Two different types of proposals have been put forward to explain a bilingual disadvantage in retention, mirroring those described above regarding children's ME performance.

The first proposal is that mono- and bilinguals apply different word learning strategies: The idea is that, based on their experience with many-to-one mappings across languages, bilinguals learn around their second birthday that ME is not a reliable word-learning strategy for them (although ME might still be used for communication in-the-moment; Kalashnikova et al., 2018). The second proposal is that bilingual children have the same strategies available as monolinguals do, but they cannot apply them for various reasons. For example, bilinguals may be too uncertain about the label of the familiar distractor objects because of their vocabulary delay in the test language or because of competition from labels for that object in their other language (Rocha-Hidalgo et al., 2021). Thus, the contrast between word knowledge of the distractor and the novel object may be sufficient for disambiguation, but not for retention (Rocha-Hidalgo et al., 2021). Bilingual children may also face further performance issues, because they may not feel comfortable enough in the test language to perform according to their potential.

To decide between these two proposals, research with young preschoolers may be especially informative. If bilinguals' experience leads them to discard ME as a basis for word learning from their second birthday onwards, then a difference between mono- and bilingual children should become more pronounced in young preschoolers. In contrast, if bilinguals have the same strategies available, but are hindered from applying them by performance issues (e.g., uncertainty about the familiar distractor label), then difficulties should not persist in preschool years (especially if highly familiar distractor objects are used).

Methodological considerations

The mixed pattern of findings regarding bilingual disadvantages may reflect methodological issues. One potential issue concerns small and unbalanced sample sizes.

Small sample sizes generally come with problems of statistical power. This becomes particularly problematic if the sample sizes are unbalanced (i.e., more mono- than bilinguals are tested) or if the reported difference in performance is established by testing each group's performance against chance, with no direct comparison of mono- and bilingual samples. Furthermore, since bilinguals, in contrast to monolinguals, are often not tested in their first language, they may feel less comfortable in the test language, causing disadvantages that are independent of the capacities of interest.

In sum, past research suggested bilingual disadvantages in disambiguating novel word meanings in the classic ME task and in retaining these mappings subsequently. However, mixed findings and methodological concerns highlight the need for preregistered research that tests balanced samples of mono- and bilingual preschoolers, based on *a priori* power analyses, to assess the robustness of the suggested effects. Given the suggestion that the bilingual disadvantage in novel word retention develops at age 2 (see Kalashnikova et al., 2018), the assessment of young preschoolers is especially relevant.

Bilingual Pragmatic Advantages

A second, more or less independent, line of research has suggested that bilingual have *advantages* in other word learning areas. Specifically, bilinguals may outperform their monolingual peers in their pragmatic and social-communicative skills (e.g., Fan et al., 2015; Yow & Markman, 2015). Again, there are different explanations for the potential differences between mono- and bilinguals.

First, bilinguals' language experience may lead to advanced communicative and socio-cognitive abilities. These advantages may be a consequence of the communicative challenges they are facing in their daily lives – such as communicative failures, misunderstandings, and adapting to an environment using different languages (Fan et al., 2015; Liberman et al., 2017; Wermelinger et al., 2017) – and may even be a way of compensating for an initial lag in word learning (Siegal et al., 2009). Bilinguals' continuous demands to flexibly adjust their linguistic interactions to their conversation partner, may train their communicative skills, as well as their perspective taking (Schroeder, 2018) and executive function in general (Ware et al., 2020).

Second, the bilingual pragmatic advantages may also be a consequence of a systematic selection bias. Bilingual populations do not only differ from monolinguals in terms of their language background. Bilingual families often immigrated from another country (and cultural background) and may potentially be more open-minded and socially-sensitive, which may in turn lead to social-cognitive advantages that are not specifically due to their language experience (see, e.g., Gampe et al., 2020 for an example in which not bilingualism *per se*, but the cultural background of child and caregiver shapes communicative interactions).

The empirical basis for the proposed bilingual pragmatic advantages is, again, mixed (van Wonderen et al., 2023). Bilingual advantages are mainly found in younger children, and may depend on the specific task or ability in question (see Antoniou et al., 2020). They may be especially pronounced in tasks that tap more basic pragmatic and social-communicative skills, such as perspective taking (Fan et al., 2015; Liberman et al., 2017) and Theory of Mind in general (Schroeder, 2018), understanding referential intent (Yow et al., 2017; Yow & Markman, 2015), or repairing communication failures (Wermelinger et al., 2017). Differences may be less robust, however, in more complex pragmatic abilities, such as irony, sarcasm, metaphors or implicatures (Antoniou et al., 2020; Syrett et al., 2017; but see Siegal et al., 2009). Overall, past research on bilingual pragmatic advantages is relatively sparse and it remains open if the proposed advantages are replicable and to which specific pragmatic abilities they apply.

The Current Study

Past research has suggested that mono- and bilingual children may differ in the strategies and pragmatic abilities underlying their word learning. However, there is uncertainty regarding the robustness, as well as the explanations for these potential differences. In the current project, we focus on two central areas in which differences have been suggested: i) a bilingual disadvantage with regard to the use of ME to disambiguate and learn novel word-object links and ii) a bilingual advantage in using social-pragmatic information for word learning.

Combining both lines of research, we tested whether mono- and bilingual children differ in the strategies and pragmatic abilities they use for referent disambiguation and word learning. We tested comparable samples of mono- and bilingual 3-year-olds and adults (as a comparison group) with the same first language (German) who participated from their homes. We assessed their disambiguation and retention performance in two conditions (within-subjects): In the classic ME condition, a novel label was used in the presence of one novel and one familiar object. In the pragmatic condition, we presented two novel objects and the referent of the novel word could be pragmatically inferred based on common ground information (discourse novelty; see, e.g., Bleijlevens et al., 2023; Bohn et al., 2022).

With our design, we aimed to address how robust and persistent bilingual advantages and disadvantages are in preschoolers, and how specific they are to the proposed areas. We predicted the following: First, if bilingual *disadvantages* are robust and specific to the ME context because bilinguals learn that ME is not a reliable word learning strategy for them (Kalashnikova et al., 2018), then we expected bilingual 3-year-olds to show a weaker ME effect and weaker subsequent retention, but similar (or even better) performance in the pragmatic condition. Second, if bilingual *advantages* are robust and extend to children's use of common ground information, bilingual children should outperform monolinguals in the pragmatic condition – potentially regarding both, referent disambiguation and retention. However, if differences between mono- and bilingual children are due to *performance issues* and sampling biases rather than the application of different word learning strategies, we would not expect any differences between both groups in our study with highly comparable mono- and bilingual samples.

Method

We preregistered the experimental design, procedure, sample sizes, and statistical analyses on OSF (https://osf.io/9epuw/). The complete study materials, data, analysis scripts, and details regarding the sample size calculation, the counterbalancing/randomization plan, and results are accessible on OSF as well. This project has been approved by the ethics committee of the Institute for Psychology, University of Göttingen (project number 317b).

Participants

Children

The final sample for the main analyses included 74 typically developing 3-year-old children (36-48 months, M= 40.9, SD= 3.6; 34 females, 47 males, 1 without gender indication): 37 monolingual and 37 bilingual children. All children had German as their first language. Children were included in the bilingual group if they were exposed to at least one additional language regularly by one of their parents, constituting at least 20% of their language input. Additional languages included English, Spanish, Russian, Chinese, Danish, French, Polish, Turkish, Japanese, Dutch, Italian, Arabic, Romanian, Portuguese, Czech and Bulgarian. At least one parent had to be fluent in German to assist the child in participating in the study.

Eight additional children participated but did not meet the criteria for either the mono- or bilingual group, because they had regular contact with a second language but did not receive at least 20% input in this language. These children are referred to as bilingual in a broader sense and were only included in an exploratory analysis (see Appendix C).

In addition, 10 children participated, but were excluded from any analyses because of not providing any data in test trials (1) or based on our pre-registered exclusion criteria:

technical issues (1), at least one mistake on familiarization trials (3) and uncooperative behavior (5).

We determined the sample size of 74 children *a priori* via data simulation. The simulation was based on the analysis of a difference between mono- and bilinguals retention performance in the Classic ME condition. The goal was to obtain .8 power to detect the assumed effect size. Children were invited via databases of parents who had agreed to be contacted for studies.

Adults

The final sample included 86 adults with German as their first language (19-72 years, M=36.3, SD=13.5 years; 38 females, 48 males). We determined this sample size *a priori* via data simulation, based on the analysis of an effect of bilingualism on adults' retention performance in the Pragmatic condition. Adults were recruited via an online platform (www.prolific.com) and compensated for their participation at the recommended rate (£9/h). Bilingual adults (n=43) indicated on Prolific that their first language was German, that they were raised with two or more languages and are fluent in their native language (German) as well as at least one other language. Monolinguals (n=43) indicated that their first language was German and that they were raised with their native language only. For further information regarding the participants' language backgrounds see Supplement A.

Design

We used a 2 (condition: "Classic ME" vs. "Pragmatic") x 2 ("monolingual" vs. "bilingual") factorial design with conditions being tested within-subjects. Children were tested in one test trial per condition and adults in two.

Stimuli

For auditory stimuli, we recorded three female German native speakers, one for each animal speaker in the experimental phases. Two non-words ("ergi" and "sude"), that matched German phonology, served as novel labels in the referent disambiguation trials and six German known words served as labels in practice ("apple", "house", "flower", "bus") and familiar-label trials ("ball", "shoe"). The familiar distractor for the Classic ME task was a car (and for adults additionally a flower). Based on the Wordbank (M. C. Frank et al., 2017), each of these words (those used as labels in practice and familiar-label trials, as well as the word car) is, on average, produced by 94-100% of 2.5-year-old German-speaking children and expected to be well known to all of our 3-year-olds.

For visual stimuli, we used unknown object images from the NOUN database (Horst & Hout, 2016), images provided by Bohn et al. (2022; see also Bleijlevens et al., 2023) and open source material. Videos were created via Powerpoint.

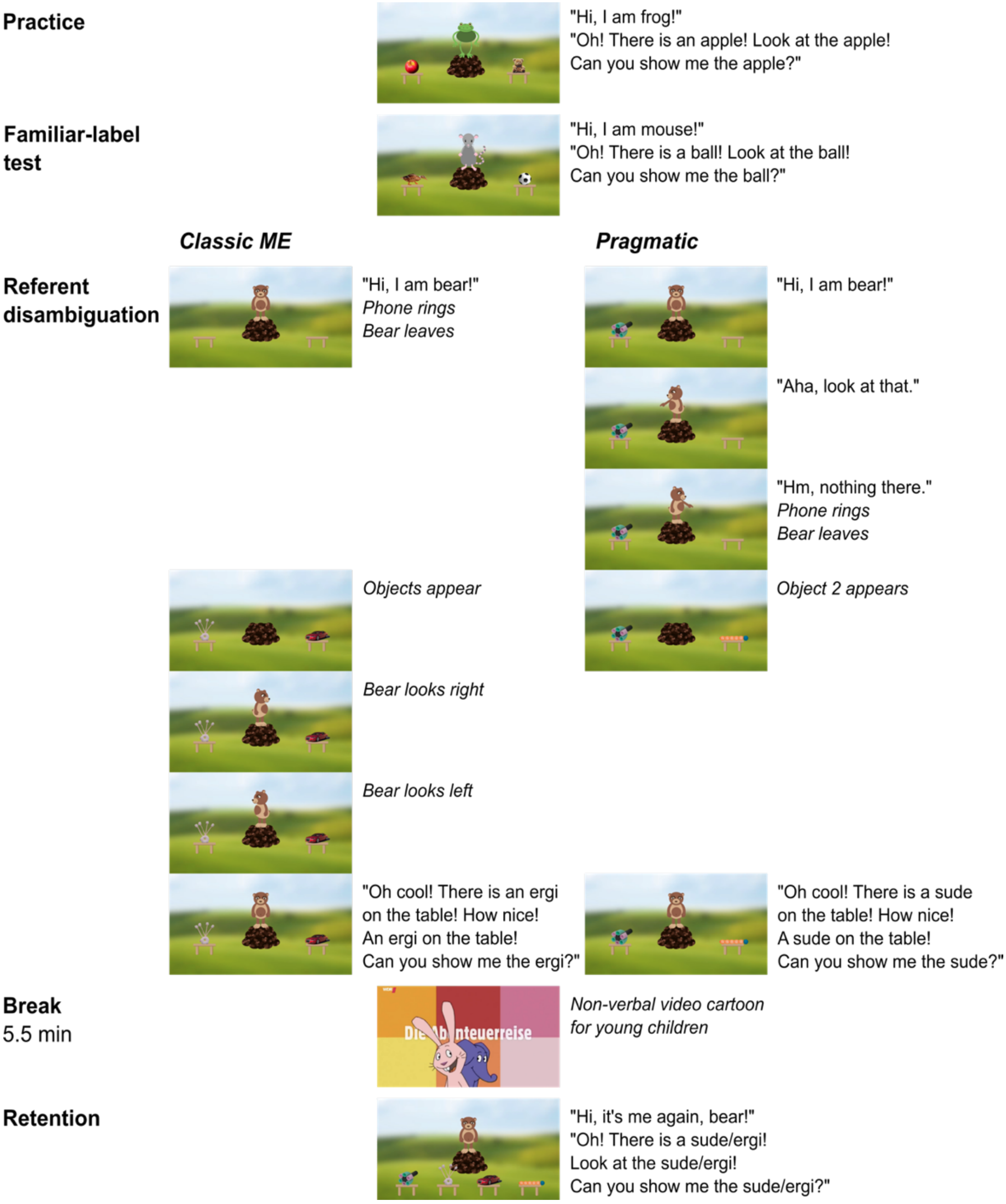
Procedure

The study was conducted online (due to the Covid-19 pandemic). The child version was a synchronous online study (using BigBlueButton video conferencing), the adult version was an asynchronous online study presented via Labvanced (Goeke et al., 2017). Participants watched short, animated videos which asked them to point to different objects. We video-recorded children's, but not adults', testing sessions. After providing informed consent, each participant was presented with 10 (children) or 12 (adults) trials in four experimental phases: Practice (children: 4 trials, adults: 2 trials), Familiar-label test (2 trials), Referent disambiguation (children: 2 trials, adults: 4 trials for) and Retention (children 2 trials, adults: 4 trials; Figure 1). We created 16 experimental versions for counterbalancing/randomization of factors in the tasks (for details see below).

On each trial, an animal speaker, located at center stage and looking straight ahead, asked for one of the presented objects. Adult participants directly clicked on the objects on the screen. Children selected objects by pointing at them. A letter then appeared underneath each of the objects on screen and parents were instructed to indicate their children's choice by reading aloud the letter displayed under the chosen object. Parents and the experimenter were not permitted to help or interfere in children's choices in any way (except in Practice trials). When children did not respond, the experimenter encouraged the participant by asking "Just choose what you think is right". After the main experiment, children participated in a German receptive vocabulary test and adults were asked to describe the selection strategies.

Figure 1

Experimental Procedure: Example Trials for Each Phase in Both Conditions



Note. The wording of the test question was hold constant across conditions, as were the main elements involved in the overall scene (e.g., bear leaving in response to a phone ringing, bear turning towards each table before asking the test question etc.).

Practice

Frog introduced herself. Then, on each trial, two known objects appeared on top of the screen and descended until they rested each on one of the two empty tables in front of frog. Frog asked the participant to point at a specific object by asking “Oh, there is a [known label]! Look at the [known label]. Can you show me the [known label]?” The experimenter interacted with the child until she picked the correct object and choices were followed by positive feedback from frog.

Practice trials were presented in fixed order (children: apple-house-flower-bus; adults: bus-house), and target locations (children: left-right-right-left; adults: left-right). Target locations were, thus, fully counterbalanced for each participant.

Familiar-Label Test

Mouse introduced herself. In each of the two familiar-label trials, two known objects fell down on the two empty tables in front of her, followed by mouse's request to show one of them to her. In contrast to the practice trials, participants did not receive any feedback. Subjects who made at least one mistake in familiar-label trials were excluded from analyses.

The object pairings (ball & duck, shoe & banana) and targets (ball, shoe) were fixed. The order of trials and target locations were randomized, either across experimental versions (for children) or in-the-moment by the experimental platform (for adults). Within participants, target locations were fully counterbalanced, with one target being presented on the left table and one on the right.

Referent Disambiguation

Bear introduced herself. The following procedure within each trial depended on the condition. Children were presented with one trial per condition in counterbalanced order, adults with two.

Classical ME condition. A phone starting ringing and bear left the scene, disappearing inside the hill. Then, two objects, one familiar and one unfamiliar object, appeared on top of the screen and descended until they rested on the tables. Bear reappeared and turned towards each table in turn. Then, bear looked straight ahead and 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]?".

Pragmatic condition. One object (the distractor) was laying on one of the two tables. Bear turned to each table. While looking and pointing at the empty table she said "Hm, nothing there", and while turning to the occupied table she said "Aha, look at that". A phone started ringing, and bear left the scene by disappearing inside the small hill. Meanwhile, a second novel object fell down and rested on the empty table, and then bear reappeared. Just as in the ME condition, bear then looked straight ahead and 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]?"

In the child study, across experimental versions, we counterbalanced the order of conditions, labels, and gaze directions (first to target/distractor side), and we randomized target locations (left/right) and the assignment of unknown objects to a role in the experiment (target or distractor in the Pragmatic task, or distractor in the Classic ME task). In the adult study, which included 4 instead of 2 referent disambiguation trials, all of these variables were randomized across versions. We fully counterbalanced target locations within participants, and gaze directions within participants and conditions.

Break

A children's non-verbal video cartoon was played, serving as a time delay of 5.5 minutes prior to retention trials.

Retention

Participants were presented with one retention trial per newly learned label (i.e., two retention trials for children and four for adults). In each trial, bear was standing behind four tables when four objects fell down onto them. Without changing her frontal gaze direction, bear said "Oh, there is a [novel label]! Look at the [novel label]! Can you show me the [novel label]?".

The four presented objects were identical in both conditions. In the child version, we presented all four objects they had encountered in the two referent disambiguation trials: both target objects and both distractor objects from the "ergi" and "sude" disambiguation trials. Across the experimental versions for children, we counterbalanced the label order (ergi/sude first), the correspondence of label order relative to the label order in referent disambiguation trials (same/different), and object locations, such that across these versions, each object was presented equally often at each position and changed its position between trials. In the adult version, we extended this approach to the four presented trials. For details see Appendix A.

Receptive Vocabulary Test (Children Only)

Following the main experiment, children participated in a German receptive vocabulary test for 3- to 8-year-olds (Bohn et al., 2023), including 20 trials.[1] On each trial, four different pictures were presented on screen and a voice asked the child to point at one of them. Note that only part of our sample (20 monolinguals, 19 bilinguals, 8 bilinguals in the broader sense) provided data on this test, because we only started implementing the test later in the data collection process and because some children failed to concentrate after the main study.

[1] Note that this was an earlier version of the test including only 20 trials (instead of 22 in the final version).

Selection Strategies (Adults Only)

At the end of the experiment, we asked adults to describe how they decided on the object they selected in each condition (one question per condition). The question was accompanied by a screenshot of the participant's first referent disambiguation trial in that condition (in counterbalanced order). We asked, e.g., "Please think back to the first game with bear: Could you shortly describe, how you decided which object 'ergi' is referring to in the situation on the picture?" and participants answered in an open text field.

Measures

Object Choices

We measured object choices for adults via their mouse clicks, and for children via their pointing gestures, confirmed by parents' reading out of the corresponding letter. We were interested in *correct choices* in referent disambiguation trials, i.e., selecting the novel (vs. known/pre-exposed) object and in *consistent choices* in retention trials, i.e., selecting the same object they had previously selected in the corresponding disambiguation trial.

Adults' Response Times

We measured adults' response times in referent disambiguation and retention trials. Response times started with the first label onset and ended with their mouse click on one object.

Adults' Selection Strategies

A blinded coder assigned adults' descriptions of their selection strategies to one of five pre-registered categories: "speaker intent", "nameability", "familiarity", "perceptual features", and "explicit guessing" (Table 1). Reliability coding by a second blinded coder for all trials revealed 80% agreement. Cases of disagreement (34 out of 162) were discussed with a third coder until a joint decision was reached. The majority of disagreements was due to either assignments to closely related categories (i.e., disagreement between nameability and familiarity (n=8) or between perceptual features and explicit guessing (n=2)), or due to one coder, but not the other, refraining from assigning any category (n=17).

Statistical Analysis

For data analysis, we used R (version 4.2.1; and RStudio (version 2023.6.0.421; Posit team, 2023). Appendix B lists all functions and packages used. The data set, R scripts, analysis results and assumption tests are accessible on OSF (https://osf.io/9epuw/). If not stated otherwise, we followed our preregistered analysis plan and the model assumptions were met.

Table 1

Categories of Adults' Specific Reasoning Strategies

Strategy	Explanation	Examples
Speaker intent	Reasoning based on the speaker, her behavior/intentions	"One object was already present and bear has seen it, but not named it. Then bear was surprised when the second object appeared – therefore it should be this one."
Nameability	Reasoning based on the nameability of an object	"The other object already has a name."
Familiarity	Reasoning based on the participant's familiarity with the object	"I am clearly familiar with one of the objects, therefore it must have been the one I didn't know."
Perceptual features	Reasoning based on objects' perceptual (visual/auditory) properties or salience	"The word seemed to fit the shape of the individual elements of the object."
Explicit guessing	Indication of own ignorance/ selection based on intuition	"purely intuitively"/ "I don't know the word toma, so I just guessed"

Note. As preregistered, we distinguished the categories "nameability" and "familiarity". However, we realized that many given answers were in line with both categories (e.g., "It can't be the car"), because they do not differentiate if the distractor was excluded based on its name or familiarity. In these cases, we decided to code "nameability" whenever the objects' name was mentioned, leading to a high number of "nameability" and a relatively low number of "familiarity" codings.

Before interpreting model parameters, we tested for the overall effect of our fixed effects for each model with more than one predictor by using Likelihood Ratio Tests comparing the fit of the full model to that of a null model, lacking the predictors of interest. This way, we avoided "cryptic multiple testing" (Forstmeier & Schielzeth, 2011).

As preregistered, we removed non-significant interactions from full models in a stepwise fashion, starting with non-significant higher-order (3-way) interactions (e.g., language background x condition x age group), and followed by non-significant lower-order

(2-way) interactions (e.g., language background x condition, language background x age group, or condition x age group, as applicable).

Object Choices in Referent Disambiguation

To test whether participants' performance in referent disambiguation trials differed between language backgrounds or conditions, we fitted a GLMM with binomial error distribution. We predicted participants' correct object choices by language background (monolingual/bilingual), condition (Classic ME/Pragmatic), age group (children/adults), and all of their possible interactions. Additionally, we added the speaker's gaze order (first to target/distractor) as a control variable and random intercepts for participants. This analysis differed from the preregistered one in that we replaced the predictor continuous age (in years) by age group. The new model eased the interpretation and description of the results, but revealed the same pattern of results as the preregistered one (see Appendix C: Figure C1 and Table C1).

Adults' Response times. To test for effects of language background and condition on adults' processing speed, we fitted a LMM. We predicted adults' log-transformed response times by language background, condition and their interaction. We added age (z-transformed)[2] and the speaker's gaze order as control variables, and random intercepts for participants.

Adults' Selection Strategies. To test adults' differential use of strategies across conditions, we fitted a multinomial mixed effects model. We predicted adults' strategies by language background, condition and their interaction, and added random intercepts for participants.

Object Choices in Retention. To test whether participants' retention performance was affected by language group and/or conditions, we fitted a GLMM with binomial error distribution. We predicted participants' consistent object choices by language group, condition, age group, and all of their possible interactions. We added random intercepts for participants. As above, we decided to ease the interpretation and communication of the results by replacing the preregistered predictor continuous age (in years) by age group. This model revealed the same pattern of results as the preregistered one (see Appendix C: Figure C2 and Table C2).

[2] We preregistered to log-transform age for this analysis. However, while this makes sense in a model including both children's and adults' data, it is not necessary in a model including only adults' data. Note that this decision did not change the results.

Exploratory Analysis

First, we tested whether mono- and bilinguals differed in their L1 (German) vocabulary size. We fitted a GLMM with binomial error distribution, predicted their correct choices in the vocabulary test (on a trial basis) by language background, and added random intercepts for participants.

Second, in all analyses described above, we included only bilingual children in the stricter sense. To test whether the inclusion of bilinguals in a broader sense changed the results, we fitted the previously described models for participants' correct choices in referent disambiguation trials (Table C3) and their consistent choices in retention trials (Table C4) again after including this broader group of bilingual children. The inclusion of these children did not change the main pattern of results (for details see Appendix C).

Third, to assess whether different measures of language background/ bilingualism may have had an effect on adults' disambiguation and retention, we exploratorily ran the main analyses described above (object choices in disambiguation and retention trials, and response times in disambiguation trials) again and replaced the main bilingualism predictor (raised bilingual) with all of our alternative measures. None of the other bilingualism measures affected adults' performance in the expected direction (Supplement A).

Results

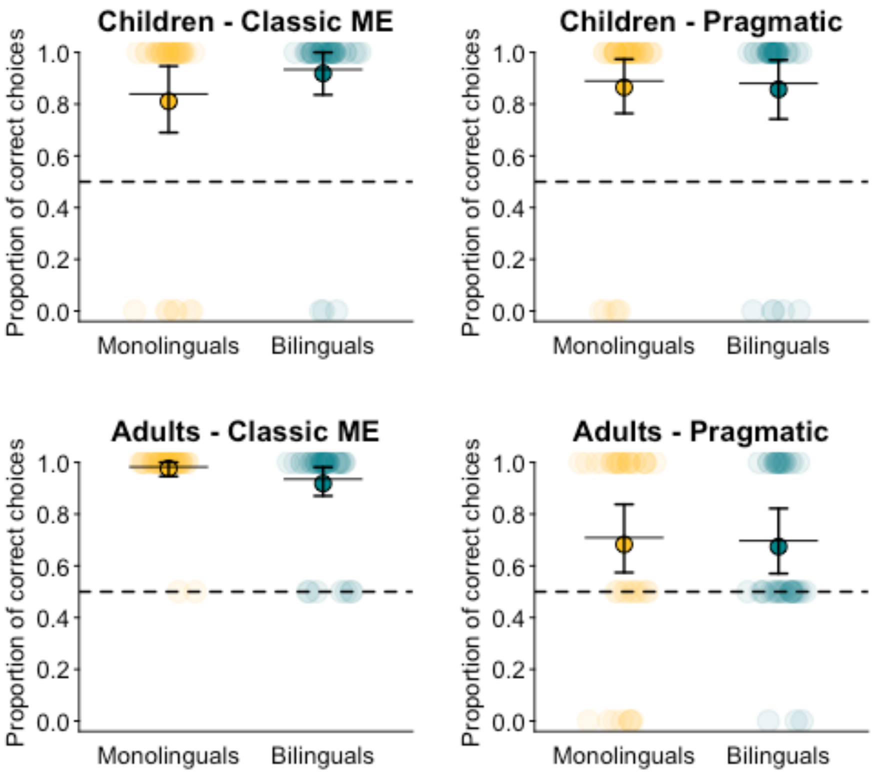
Correct Object Choices in Disambiguation Trials

Children and adults selected the target object significantly above chance level in both conditions (Figure 2: the bootstrapped confidence intervals do not include chance level).

As preregistered, in a stepwise fashion, we first removed the non-significant 3-way interaction (b = 2.38, SE = 1.37, p = .082), and then the non-significant 2-way interactions (language background x condition: b = 0.33, SE = 0.63, p = .603; language background x age group: b = -0.83, SE = 0.68, p = .225). See Table 2 for the results of the full model. The final reduced model included all main effects and only the significant interaction of condition and age group (b = -2.28, SE = 0.64, p < .001). Overall, language background did not affect the performance in disambiguation trials (b = -0.07, SE = 0.30, p = .829). Children's performance did not differ between conditions (b = -0.03, SE = 0.50, p = .956), but adults performed significantly better in the Classic ME than in the Pragmatic condition (b = -2.31, SE = 0.41, p < .001). In fact, adults performed significantly better than children in the Classic ME condition (b = 1.10, SE = 0.51, p = .033), but significantly worse than children in the Pragmatic condition (b = -1.19, SE = 0.43, p = .005). The speaker's gaze order did not affect performance (b = -0.38, SE = 0.28, p = .168). The reduced model described the data significantly better than the corresponding null model ($\chi^2(4) = 48.03$, $p < .001$) and did not differ significantly from the full model including all interactions ($\chi^2(3) = 4.74$, $p = .192$).

Figure 2

Correct Object Choices in Referent Disambiguation Trials



Note. Transparent dots represent the proportions of correct object choices per participant, based on those trials in which any choice was made (children: $n_{\text{monolingual}} = 74$, $n_{\text{bilingual}} = 72$; adults: $n_{\text{monolingual}} = 167$, $n_{\text{bilingual}} = 172$) and circled dots the aggregated proportions per group and condition. Horizontal lines indicate the predicted probabilities; and vertical lines the 95% confidence intervals, both obtained by the GLMM and calculated via bootstrapping with 1000 boots.

Table 2

Model Predicting Correct Choices in Referent Disambiguation Trials by Language Background, Condition, Age Group, Their Interactions, and Gaze Order

	Estimate	*SE*	*p*	95% *CI*
Reference groups: monolinguals, Classic ME condition, children				
Intercept	1.85	0.52	<.001	1.01, 3.08
Language background	1.00	0.78	.201	-0.45, 10.44
Condition	0.43	0.66	.513	-0.90, 2.03
Age group	2.37	0.87	.006	0.88, 11.74
Gaze order	-0.37	0.28	.180	-0.93, 0.16
Language background x condition	-1.09	1.03	.292	-10.32, 0.95
Language background x age group	-2.35	1.15	.042	-12.33, -0.18
Condition x age group	-3.57	1.03	<.001	-13.03, -1.91
Language background x condition x age group	2.38	1.37	.082	-0.12, 12.78
Reference groups: bilinguals, Pragmatic condition, adults				
Intercept	1.03	0.33	.002	0.46, 1.72
Language background	0.06	0.40	.884	-0.70, 0.84
Condition	1.84	0.48	<.001	1.04, 3.08
Age group	1.17	0.59	.049	0.15, 2.67
Gaze order	-0.37	0.28	.180	-0.93, 0.16
Language background x condition	1.29	0.90	.150	-0.22, 11.91
Language background x age group	0.03	0.83	.968	-1.73, 1.82
Condition x age group	-1.19	0.92	.197	-3.16, 9.15
Language background x condition x age group	-2.38	1.37	.082	-13.81, 0.12

Note. GLMM with binomial error distribution on participants' correct choices in referent disambiguation trials with language background (monolingual/bilingual), condition (Classic ME/Pragmatic), age group (children/adults) and all of their interactions as predictors, gaze order (first to target/distractor) as control variable and random intercepts for participants (*SD* = 0.80). $N_{observations}$= 485. N_{groups} = 160. Confidence intervals were obtained via bootstrapping with 1000 boots.

Adults' Response Times

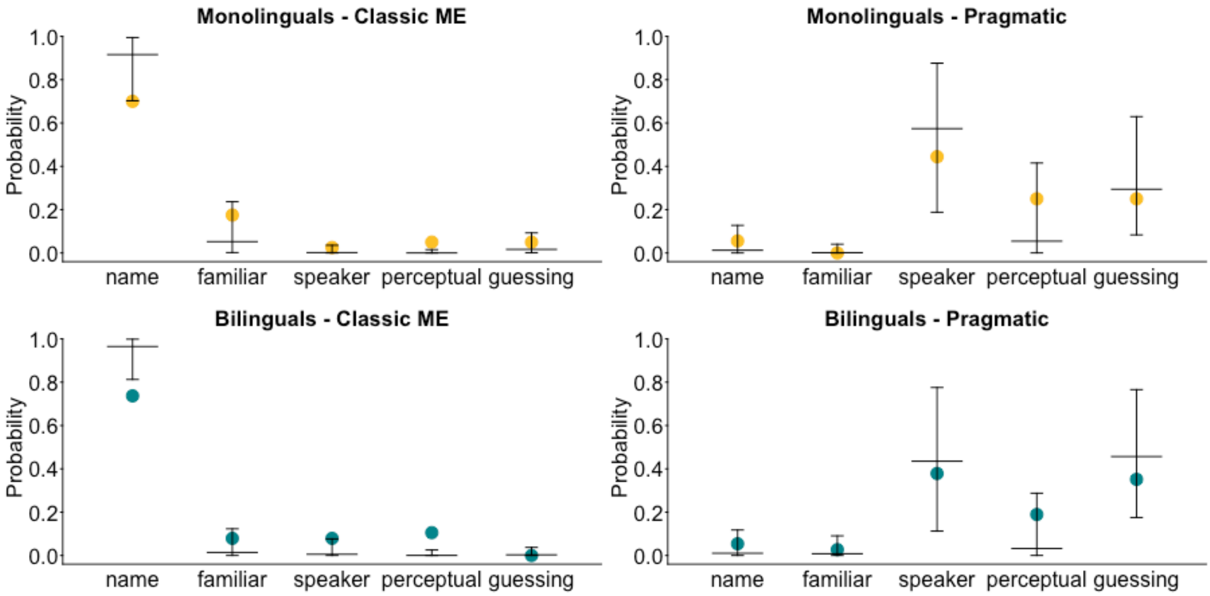
Since the full model did not reveal a significant interaction between language background and condition (*b* = 0.03, *SE* = 0.08, *p* = .710), we interpreted the reduced model lacking this interaction. There was no significant effect of language background on adults' response times (*b* = -0.05, *SE* = 0.09, *p* = .560). However, in line with their object selections, adults responded significantly faster in the Classic ME (*M* = 7.6 sec, *SD* = 3.7 sec) than the Pragmatic condition (*M* = 8.8 sec, *SD* = 4.0 sec; *b* = 0.18, *SE* = 0.04, $p < .001$). There were no significant effects of adults' age (*b* = 0.06, *SE* = 0.05, *p* = .159) nor the speaker's gaze order (*b* = -0.03, *SE* = 0.04, *p* = .491) on their response times. The reduced model described the data significantly better than the corresponding null model ($\chi^2(2) = 20.11$, $p < .001$) and did not differ significantly from the full model including the interaction ($\chi^2(1) = 0.14$, $p = .710$).

Adults' Selection Strategies

Mono- and bilinguals did not differ in the strategies they used to disambiguate words in our tasks (Figure 3: confidence intervals of one language group do not include fitted values of the other). In the Classic ME condition, both mono- and bilingual adults described strategies in line with the nameability category significantly more often than any other category. In the Pragmatic condition, strategies based on the speaker's intentions were the most prevalent of the five categories (described in 41% of the Pragmatic trials) for both mono- and bilinguals (Table 3). These "speaker intent" strategies were significantly more frequent than strategies based on the objects' nameability or familiarity, and significantly more frequent than selections based on perceptual features. However, the number of adults who decided on an object by guessing and/or based on its perceptual features was still unexpectedly high.

Figure 3

Adults' Reasoning Strategies for Referent Disambiguation



Note. Dots show the proportions of actual selection strategies for those trials in which adults indicated a strategy matching any of the five preregistered categories (monolinguals: $n_{ClassicME} = 40$, $n_{Pragmatic} = 36$; bilinguals: $n_{ClassicME} = 38$, $n_{Pragmatic} = 37$; 21 of the 172 trials included answers that did not match one of our categories). Horizontal lines indicate the predicted probability of this strategy by the multinomial mixed model and vertical lines their 95% confidence intervals.

Table 3

Adults' Reasoning Strategies for Referent Disambiguation

Strategy	Classic ME		Pragmatic	
	Monolinguals	Bilinguals	Monolinguals	Bilinguals
Speaker intent	1 (2.5%)	3 (7.9%)	16 (44.4%)	14 (37.8%)
Nameability	28 (70.0%)	28 (73.7%)	2 (5.6%)	2 (5.4%)
Familiarity	7 (17.5%)	3 (7.9%)	-	1 (2.7%)
Perceptual features	2 (5.0%)	4 (10.5%)	9 (25.0%)	7 (18.9%)
Explicit guessing	2 (5.0%)	-	9 (25.0%)	13 (35.1%)

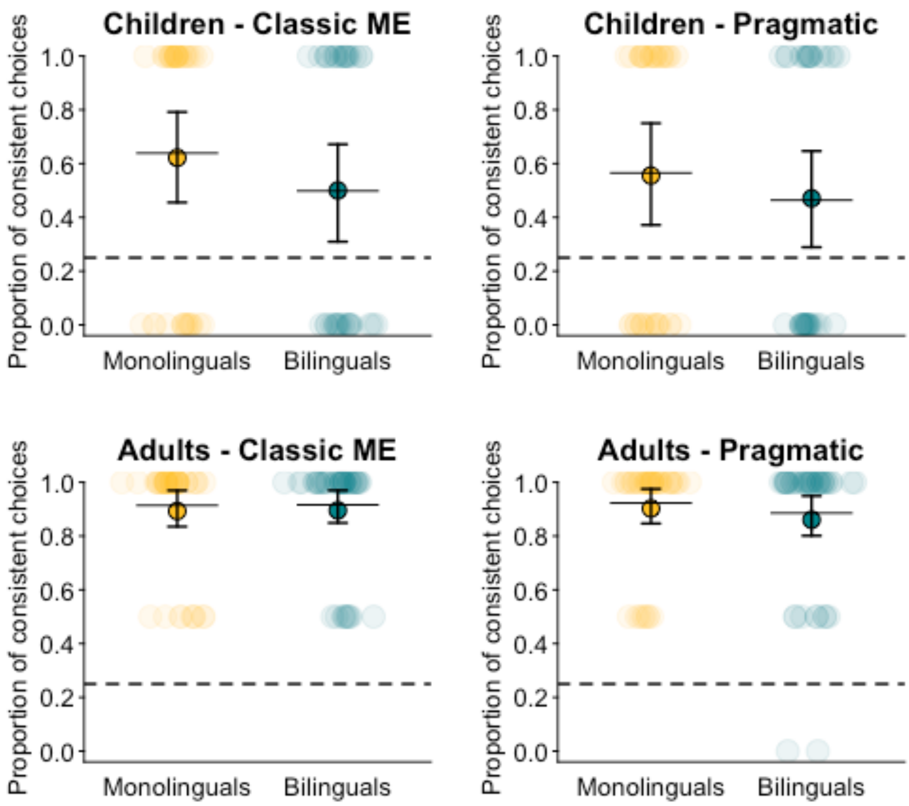
Note. Percentage of coded strategies per language group and condition for those responses that were codable (n = 151 out of 178).

Object Choices in Retention Trials

Children and adults, both mono- and bilinguals, made consistent choices in retention trials above chance in both conditions, as confirmed by the bootstrapped confidence intervals (Figure 4: confidence intervals do not include the chance level of 0.25).

Figure 4

Consistent Object Choices in Retention Trials



Note. Transparent dots represent the proportions of consistent object choices per participant, based on those trials in which any choice was made (children: $n_{monolingual}$ = 73, $n_{bilingual}$ = 70; adults: $n_{monolingual}$ = 166, $n_{bilingual}$ = 172) and circled dots the aggregated proportions per group and condition. Horizontal lines indicate the predicted probabilities; and vertical lines the 95% confidence intervals, both obtained by the GLMM and calculated via bootstrapping with 1000 boots.

The full model did not reveal any significant interactions (Table 4). As preregistered, we removed non-significant interactions from the model in a step-wise manner.[3] The resulting

[3] In the first step, we removed the non-significant 3-way-interaction of language background x condition x age group (b = -0.63, SE = 1.02, p = .535). In the second step (in the resulting reduced model), we

reduced model including only main effects showed that, although adults had to remember 4 word-object-links and children only 2, adults' performance in retention trials was significantly better than children's (*b* = 2.15, *SE* = 0.32, *p* = <.001). However, there were no significant effects of language background (*b* = -0.35, *SE* = 0.29, *p* = .223) or condition (*b* = -0.18, *SE* = 0.25, *p* = .483). The reduced model described the data significantly better than the corresponding null model ($\chi^2(3) = 56.7$, $p < .001$) and did not differ significantly from the full model including all interactions ($\chi^2(4) = 0.75$, $p = .945$).

Table 4

Model Predicting Consistent Choices in Retention Trials by Language Background, Condition, Age Group, and Their Interactions

	Estimate	*SE*	*p*	95% *CI*
Reference groups: monolinguals, Classic ME condition, children				
Intercept	0.57	0.39	.142	-0.18, 1.34
Language background	-0.58	0.55	.290	-1.69, 0.43
Condition	-0.31	0.51	.547	-1.39, 0.77
Age group	1.79	0.56	.001	0.76, 3.10
Language background x condition	0.17	0.73	.815	-1.28, 1.61
Language background x age group	0.61	0.77	.432	-1.01, 2.15
Condition x age group	0.42	0.73	.565	-1.25, 1.97
Language background x condition x age group	-0.63	1.02	.535	-2.67, 1.59
Reference groups: bilinguals, Pragmatic condition, adults				
Intercept	2.04	0.38	< .001	1.39, 2.92
Language background	0.43	0.53	.415	-0.63, 1.59
Condition	0.35	0.48	.473	-0.70, 1.35
Age group	-2.19	0.55	< .001	-3.41, -1.22
Language background x condition	-0.46	0.72	.518	-1.85, 0.98
Language background x age group	-0.03	0.77	.972	-1.69, 1.66
Condition x age group	-0.21	0.71	.767	-1.75, 1.26
Language background x condition x age group	0.63	1.02	.535	-1.59, 2.65

removed all non-significant 2-way interactions, i.e., condition x language background (*b* = -0.15, *SE* = 0.51, *p* = .765), condition x age group (*b* = 0.10, *SE* = 0.51, *p* = .852), and language background x age group (*b* = 0.29, *SE* = 0.57, *p* = .616).

Note. GLMM with binomial error distribution on participants' consistent choices in retention trials with language background (monolingual/bilingual), condition, age group (children/adults) and all of their interactions as predictors, and random intercepts for participants (*SD* = 0.79). $N_{observations}$= 481. N_{groups} = 159. Confidence intervals were obtained via bootstrapping with 1000 boots.

Vocabulary Size (Exploratory)

Bilingual children in our sample had a significantly lower vocabulary size (*M* = 7.1, *SD* = 2.8) than the monolinguals (*M* = 11.0, *SD* = 2.3; *b* = -0.81, *SE* = 0.18, $p < .001$).

Discussion

In the current study, we tested whether mono- and bilinguals differ in the strategies and socio-pragmatic skills underlying their word learning. We found that mono- and bilingual children (and adults) were equally successful at disambiguating and retaining novel word meanings in a classic ME task as well as a pragmatic task that required taking common ground into account. In contrast to prior suggestions, in our comparable samples of children with different language backgrounds, mono- and bilinguals did not differ in the strategies and pragmatic abilities they relied on to learn the meanings of novel words.

(No) Bilingual Disadvantages in Referent Disambiguation and Retention

Past research suggested bilingual disadvantages in using ME to disambiguate novel word meanings and subsequently retain these labels (Kalashnikova et al., 2018). Here, we tested mono- and bilingual 3-year-olds' (and adults') disambiguation and retention of novel words in the classic ME task. Both groups were successful at inferring that the novel label referred to the novel object and retained the labels after a short delay – without any differences between language groups. Furthermore, both groups of children performed just as successfully in the ME task as in the pragmatic task (a task that did not include any known labels or objects and in which the correct referent could be inferred pragmatically).

Our findings contradict the idea that bilinguals' language experience, particularly their experience with cross-language synonyms (i.e., many-to-one mappings) results in the development of different word learning strategies: It has been proposed that bilingual children may learn around their second birthday that ME as a word learning strategy is not reliable for them (Kalashnikova et al., 2018). However, in that case, we would have expected a) that the differences between mono- and bilinguals' performance in the ME task are especially pronounced in preschoolers and b) that bilinguals perform worse in the ME than the pragmatic disambiguation task, since any bilingual disadvantages should have been specific to tasks involving word-known distractors. In contrast, neither children's language background nor the experimental conditions affected their disambiguation and retention performance: Mono- and bilinguals performed equally well in disambiguating and learning novel word meanings, both in the ME task and the pragmatic task.

The results are in line with the idea that mono- and bilingual children have the same word learning strategies available, but were not always able to apply them successfully in previous studies for various reasons. First, bilinguals' early vocabulary delay in the test language may have led to a weaker ME effect in bilingual infants, while the 3-year-olds tested here already caught up enough to use ME to a similar extent as their monolingual peers. Second, bilingual children may not always feel comfortable in the test language, causing general performance deficits that are not due to their bilingualism per se. To avoid this confound, (a) we tested comparable samples of mono- and bilinguals who had the same first language (German), (b) children had at least one parent who was fluent in that language and also present during the test session, and (c) children were tested at home in their familiar environment. Third, methodological issues may have influenced the pattern of results in the literature, including small and unbalanced samples whose performance was not always directly compared, but separately tested against a chance value. This potential overestimation of differences was prohibited here by testing a balanced and bigger sample and running a preregistered analysis that included a direct test of the effect of bilingualism.

Importantly, the fact that we did not find bilinguals to perform any different from monolinguals in our study was probably not because bilinguals tested here were "not bilingual enough": The bilingual sample received at least 20% input in their additional language(s) which was provided daily by one of their parents. Additionally, just like in previous studies, we found bilinguals to have a smaller vocabulary in their first (and test) language. Nevertheless, there were no differences between mono- and bilingual's disambiguation and retention performance.

In sum, our study does not reveal any bilingual disadvantages in using ME for word disambiguation and learning. Our comparable samples of mono- and bilingual 3-year-olds inferred and retained novel word-object mappings in a ME task to the same extent. Their language background ultimately did not influence the strategies available to learn novel word meanings.

(No) Bilingual Pragmatic Advantages in Referent Disambiguation and Retention

Another line of past research suggested advantages of bilingual children in social-pragmatic and communicative skills that may be beneficial for understanding referential intent (e.g., Siegal et al., 2009; Wermelinger et al., 2017; Yow & Markman, 2015). Here, we tested participants in a pragmatic condition in which the correct referent could be inferred pragmatically by considering common ground (specifically, discourse novelty): Since one object was already given in the common ground, the speaker's later excitement while using a novel label indicated that this label rather referred to the novel object. The data revealed that mono- and bilinguals (both children and adults) were similarly successful at making this pragmatic inference and retaining the new word-object mapping after a delay.

While there was no difference in performance between mono- and bilinguals, we observed an unexpected difference between the two age groups: While children performed very well in our pragmatic disambiguation task, a subgroup of the tested adults showed difficulties in interpreting the pragmatic context. Their reduced performance in the pragmatic disambiguation task, as well as their strategy descriptions, indicated that these adults did not consider the common ground information, but instead guessed and/or selected an object based on its perceptual features. In contrast to the children, adults participated in an asynchronous study with no video-record. Thus, some adult participants included in the final sample may have been inattentive, thereby missing crucial elements of the pragmatic context (i.e., that bear had already looked and commented on one of the objects) – especially as during these elements participants were just meant to watch and not required to respond. Note that in previous work, adults had showed much better performance in almost identical tasks (Bleijlevens et al., 2023; Bohn et al., 2022) except that here the crucial elements were presented twice[4] thereby reducing the chance that participants might miss them due to inattention.

In neither age group, however, was there any evidence for a bilingual pragmatic advantage in using common ground to disambiguate or learn novel word meanings. How does this finding fit in the picture drawn by previous research?

[4] To achieve maximal comparability to the ME condition, we reduced the length of the pragmatic condition compared to the scene used in Bohn et al. (2022) and Bleijlevens et al. (2023). While in those studies, the speaker left and entered the scene twice to comment on both the empty table and the "old" object in the scene, she did so only once in our task. Thus, in our task the contrast in the speaker's excitement between the old and the new object was less pronounced, potentially leading to a more subtle pragmatic context. The procedural changes inadvertently resulted in a task more susceptible to be negatively affected by brief moments of participant inattention. In previous work using the longer original scenes adults' participant showed much better performance regarding both in their referent disambiguation and in their explanation of the strategies used.

There are different explanations for the current pattern of results. First, bilinguals may not actually possess any pragmatic advantages (see also Antoniou et al., 2020; van Wonderen et al., 2023). Contrary to some proposals in the literature, communicative challenges of bilinguals may not lead to improved pragmatic skills compared to monolinguals. If true, positive findings from other studies may present sampling and performance issues.

Second, bilinguals may possess pragmatic advantages that we failed to detect in our task due to ceiling effects (given that monolinguals already showed a quite high performance in our pragmatic disambiguation task). Future research should investigate the performance of mono- and bilingual children in a more demanding pragmatic word-learning task that may detect differences even in high levels of performance.

Third, bilingual pragmatic advantages may exist, but be more specific either to certain circumstances or to other areas of pragmatic skills. For example, the enhancement of pragmatic skills may only manifest in populations who are confronted with communicative challenges to a stronger extent. Given that in our sample, one parent was fluent in each of the child's languages, respectively, children may not have faced too challenging experiences that would result in the need to focus on different social cues to understand their communication partner. Alternatively, it is possible that the challenging communicative experiences of bilingual children train, for example, their ability to take their interlocutor's perspective or affect their weighing of social cues in ambiguous contexts (Fan et al., 2015; Liberman et al., 2017; Yow et al., 2017; Yow & Markman, 2015). However, they may show no effect on the fundamental skills that are underlying every child's word learning (regardless of language background), such as the ability to consider common ground information during discourse (see e.g., Bleijlevens et al., 2023; Liebal et al., 2009; Matthews et al., 2006).

To conclude, children (and adults too) need socio-pragmatic understanding in many areas of their lives, including the area of language acquisition. Irrespective of their language background, they can use these general pragmatic skills to understand the behavior and communication of others and to learn the words of our language(s). Our study revealed no advantage for bilingual children and adults in using common ground to identify the referents of novel words and retaining them. This seems plausible given that interpreting the pragmatic context is so crucial for children's social lives, regardless of whether they learn one or many languages. More research is needed to determine if bilingual pragmatic advantages do not exist at all or apply to specific pragmatic areas, e.g., those in which understanding the referential intent requires correct weighing of different cues.

The Mechanisms Behind Children's Referent Disambiguation

There is a long-standing theoretical debate focusing on how young children succeed in disambiguating novel word meanings in tasks such as the classic ME task. Three theoretical approaches have been put forward: While *lexical accounts* propose that children rely on lexical constraints such as the ME bias (assuming concepts to have only one name, e.g., "the car cannot have two names"; Golinkoff et al., 1994; Markman & Wachtel, 1988), *socio-pragmatic accounts* assume children to use their general socio-cognitive abilities to interpret the speaker's intentions (e.g., "if she meant the car, she would have used the conventional word"; Clark, 2015; Diesendruck & Markson, 2001; Tomasello, 2010), and *associative accounts* explain children's behavior by associative processes such as the attraction to novelty (Mather & Plunkett, 2012; Smith, 2000; but see Bleijlevens et al., 2023; Bleijlevens & Behne, in press). How can our findings from the classic ME task add to the debate about the mechanisms behind children's referent disambiguation?

Since bilingualism and experiential effects in general were mostly not explicated in the initial formulations of each theoretical proposal (see e.g., Markman et al., 2003), our data cannot provide clear evidence for or against certain theoretical approaches. However, some researchers made specific predictions about the role of linguistic experience that are not in line with our findings. One version of the lexical constraint accounts claims that children's development of the ME constraint is based on their experience with one-to-one mappings. According to this idea, bilingual children should not acquire the ME principle because they learn more than one word per concept (e.g., Byers-Heinlein & Werker, 2009; Houston-Price et al., 2010), at least across languages. Similarly, associative network accounts could predict a poorer ME performance for cases in which concepts have more than one label, because the associative network is shaped by language experience and the structure of bilinguals' lexicon is not sufficient to use ME (McMurray et al., 2012).

The illustrated predictions by both lexical and associative accounts are not in line with our findings. However, our data may be compatible with a specific version of these accounts in which bilinguals separate their languages and ME is only applied within a language. This is in line with studies showing that mono- and bilinguals use ME only within and not across languages (Au & Glusman, 1990) and studies showing a weakened ME effect when the novel target word is presented in isolation versus embedded in a carrier phrase that provides additional information regarding the word's language affiliation (Rochanavibhata et al., 2022). This discussion ultimately leads to the question of how and when bilingual children separate their languages (see, e.g., Byers-Heinlein, 2014).

Finally, the pragmatic word learning accounts would predict to observe a ME effect whenever children can assume that the speaker knows the conventional word for the known (distractor) object and would use it if she wanted to refer to it (Clark, 2015; Diesendruck, 2005). Socio-pragmatic accounts would therefore not predict any differences between mono- and bilingual children in our design in which the speaker was monolingual and thus knowledgeable regarding the distractor labels. The pragmatic account is thus the only one which is unconditionally supported by our data.

Conclusion

The current study investigated whether bilingual children's language experience results in differences regarding the strategies and pragmatic skills underlying their word learning. In contrast to prior suggestions, bilingual children were not less likely than their monolingual peers to disambiguate or retain novel words in a classic ME task. Their language background did not prevent them from developing or maintaining the ME strategy to learn novel words. Additionally, they were not more successful than monolinguals in using pragmatic common ground information to disambiguate and retain novel word-object mappings. Bilinguals' language experience and potential communicative challenges did not result in advanced pragmatic abilities that underly their word learning.

Our findings suggest that comparable samples of mono- and bilingual children seem to develop the same strategies and pragmatic abilities to disambiguate and learn the meanings of novel words: They can make use of their lexical knowledge and socio-cognitive understanding to infer the meanings of novel words in their language – regardless of whether they learn only one or several languages.

References

Antoniou, K., Veenstra, A., Kissine, M., & Katsos, N. (2020). How does childhood bilingualism and bi-dialectalism affect the interpretation and processing of pragmatic meanings? *Bilingualism: Language and Cognition*, *23*(1), 186–203. https://doi.org/10.1017/S1366728918001189

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

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

Bialystok, E., Luk, G., Peets, K. F., & Yang, S. (2010). Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism: Language and Cognition*, *13*(4), 525–531. https://doi.org/10.1017/S1366728909990423

Bleijlevens, N., & Behne, T. (in press). Young children and adults use reasoning by exclusion rather than attraction to novelty to disambiguate novel word meanings. *Developmental Psychology*.

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

Bohn, M., Prein, J., Koch, T., Bee, R. M., Delikaya, B., Haun, D., & Gagarina, N. (2023). oREV: An item response theory-based open receptive vocabulary task for 3- to 8-year-old children. *Behavior Research Methods*. https://doi.org/10.3758/s13428-023-02169-3

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*. Advance online publication. https://doi.org/10.1037/xge0001216

Bürkner, P.-C. (2017). brms: An R package for bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*, 1–28. https://doi.org/10.18637/jss.v080.i01

Byers-Heinlein, K. (2014). Languages as categories: Reframing the "one language or two" question in early bilingual development. *Language Learning*, *64*(s2), 184–201. https://doi.org/10.1111/lang.12055

Byers-Heinlein, K., Chen, K. H., & Xu, F. (2014). Surmounting the Tower of Babel:

Monolingual and bilingual 2-year-olds' understanding of the nature of foreign language words. *Journal of Experimental Child Psychology*, *119*, 87–100. https://doi.org/10.1016/j.jecp.2013.09.011

Byers-Heinlein, K., & Werker, J. F. (2009). Monolingual, bilingual, trilingual: Infants' language experience influences the development of a word-learning heuristic. *Developmental Science*, *12*(5), 815–823. https://doi.org/10.1111/j.1467-7687.2009.00902.x

Byers-Heinlein, K., & Werker, J. F. (2013). Lexicon structure and the disambiguation of novel words: Evidence from bilingual infants. *Cognition*, *128*(3), 407–416. https://doi.org/10.1016/j.cognition.2013.05.010

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

Davidson, D., Jergovic, D., Imami, Z., & Theodos, V. (1997). Monolingual and bilingual children's use of the mutual exclusivity constraint. *Journal of Child Language*, *24*(1), 3–24. https://doi.org/10.1017/S0305000996002917

Diesendruck, G. (2005). The principles of conventionality and contrast in word learning: An empirical examination. *Developmental Psychology*, *41*(3), 451.

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

Fan, S. P., Liberman, Z., Keysar, B., & Kinzler, K. D. (2015). The exposure advantage: Early exposure to a multilingual environment promotes effective communication. *Psychological Science*, *26*(7), 1090–1097. https://doi.org/10.1177/0956797615574699

FitzJohn, R. (2017). *ids: Generate random identifiers*. https://CRAN.R-project.org/package=ids

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, I., & Poulin-Dubois, D. (2002). Young monolingual and bilingual children's responses to violation of the mutual exclusivity principle. *International Journal of Bilingualism*, *6*(2), 125–146. https://doi.org/10.1177/13670069020060020201

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

Gampe, A., Hartmann, L., & Daum, M. M. (2020). Dynamic interaction patterns of monolingual and bilingual infants with their parents. *Journal of Child Language*, *47*(1), 45–63. https://doi.org/10.1017/S0305000919000631

Goeke, C., Finger, H., Diekamp, D., Standvoss, K., & König, P. (2017). *LabVanced: A unified javascript framework for online studies*.

Golinkoff, R. M., Mervis, C. B., & 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

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

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

Horst, J. S., & Hout, M. C. (2016). The Novel Object and Unusual Name (NOUN) Database: A collection of novel images for use in experimental research. *Behavior Research Methods*, *48*(4), 1393–1409. https://doi.org/10.3758/s13428-015-0647-3

Houston-Price, C., Caloghiris, Z., & Raviglione, E. (2010). Language Experience Shapes the Development of the Mutual Exclusivity Bias. *Infancy*, *15*(2), 125–150. https://doi.org/10.1111/j.1532-7078.2009.00009.x

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. (2015). The effects of linguistic experience on the flexible use of mutual exclusivity in word learning. *Bilingualism: Language and Cognition*, *18*(4), 626–638. https://doi.org/10.1017/S1366728914000364

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

Liberman, Z., Woodward, A. L., Keysar, B., & Kinzler, K. D. (2017). Exposure to multiple languages enhances communication skills in infancy. *Developmental Science*, *20*(1), e12420. https://doi.org/10.1111/desc.12420

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

Markman, E. M., Wasow, J. L., & Hansen, M. B. (2003). Use of the mutual exclusivity assumption by young word learners. *Cognitive Psychology*, *47*(3), 241–275. https://doi.org/10.1016/S0010-0285(03)00034-3

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

Matthews, D., Lieven, E., Theakston, A., & Tomasello, M. (2006). The effect of perceptual availability and prior discourse on young children's use of referring expressions. *Applied Psycholinguistics*, *27*(3), 403–422. https://doi.org/10.1017/S0142716406060334

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

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

Posit team. (2023). *RStudio: Integrated Development Environment for R*. Posit Software, PBC. http://www.posit.co/

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

Repnik, K. M., Chondrogianni, V., & Sorace, A. (2021). Linking disambiguation and retention in a developmental eye-tracking study with monolingual and multilingual children. *Journal of Experimental Child Psychology*, *206*, 105072. https://doi.org/10.1016/j.jecp.2020.105072

Rocha-Hidalgo, J., Feller, M., Blanchfield, O. A., Kucker, S. C., & Barr, R. F. (2021). Patterns of mutual exclusivity and retention: A study of monolingual and bilingual 2-year-olds. *Infancy*, *26*(6), 1011–1036. https://doi.org/10.1111/infa.12432

Rochanavibhata, S., Atagi, N., Schonberg, C., & Sandhofer, C. M. (2022). The role of syntactic cues in monolingual and bilingual two-year-olds' novel word disambiguation. *Infant Behavior and Development*, *68*, 101753. https://doi.org/10.1016/j.infbeh.2022.101753

Schroeder, S. R. (2018). Do bilinguals have an advantage in Theory of Mind? A meta-analysis. *Frontiers in Communication*, *3*. https://www.frontiersin.org/articles/10.3389/fcomm.2018.00036

Siegal, M., Iozzi, L., & Surian, L. (2009). Bilingualism and conversational understanding in young children. *Cognition*, *110*(1), 115–122. https://doi.org/10.1016/j.cognition.2008.11.002

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, *Becoming a word learner: A debate on lexical acquisition* (pp. 51–80). Oxford University Press.

Syrett, K., Lingwall, A., Perez-Cortes, S., Austin, J., Sánchez, L., Baker, H., Germak, C., & Arias-Amaya, A. (2017). Differences between Spanish monolingual and Spanish-English bilingual children in their calculation of entailment-based scalar implicatures. *Glossa: A Journal of General Linguistics*, *2*(1), Article 1. https://doi.org/10.5334/gjgl.76

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

van Wonderen, E., Mulder, K., Rispens, J., & Verhagen, J. (2023). Learning how to communicate: Does exposure to multiple languages promote children's pragmatic abilities? A meta-analytic review. *Cognitive Development*, *68*, 101384. https://doi.org/10.1016/j.cogdev.2023.101384

Ware, A. T., Kirkovski, M., & Lum, J. A. G. (2020). Meta-analysis reveals a bilingual advantage that is dependent on task and age. *Frontiers in Psychology*, *11*. https://www.frontiersin.org/articles/10.3389/fpsyg.2020.01458

Weatherhead, D., Arredondo, M. M., Nácar Garcia, L., & Werker, J. F. (2021). The role of audiovisual speech in fast-mapping and novel word retention in monolingual and bilingual 24-month-olds. *Brain Sciences*, *11*(1), Article 1. https://doi.org/10.3390/brainsci11010114

Wermelinger, S., Gampe, A., & Daum, M. M. (2017). Bilingual toddlers have advanced abilities to repair communication failure. *Journal of Experimental Child Psychology*, *155*, 84–94. https://doi.org/10.1016/j.jecp.2016.11.005

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

Yow, W. Q., Li, X., Lam, S., Gliga, T., Chong, Y. S., Kwek, K., & Broekman, B. F. P. (2017). A bilingual advantage in 54-month-olds' use of referential cues in fast mapping. *Developmental Science*, *20*(1), e12482. https://doi.org/10.1111/desc.12482

Yow, W. Q., & Markman, E. M. (2015). A bilingual advantage in how children integrate multiple cues to understand a speaker's referential intent. *Bilingualism: Language and Cognition*, *18*(3), 391–399. https://doi.org/10.1017/S1366728914000133

Appendix A

Counterbalancing Information: Retention Trials in the Adult Version

For the adult version of the experiment, we extended the counterbalancing approach of the child version to the higher number of test trials (four in adults vs. two in children). For the retention trials in the adult version, we presented the target and distractor object from a Classic ME disambiguation trial and the target and distractor object from a Pragmatic disambiguation trial. Therefore, we treated the "ergi & sude" and "modi & toma" referent disambiguation trials as groups (one member of them always being in the Classic ME condition and one in the Pragmatic condition) whose objects would be presented together in the retention trials. That is, e.g., in an "ergi" or "sude" retention trial, we presented both targets and distractors from their corresponding referent disambiguation trials (i.e., the same objects as for children); and the same applied to the "modi" or "toma" retention trials. Across the 16 experimental versions for adults, we randomized the order of labels, the target locations, and the locations of the other three presented objects. We made sure that each object changed its position from one trial to the next and that within adults, the target locations were fully counterbalanced, such that participants were presented with one target in each of the four locations.

Appendix B

Packages and Functions

For data handling, preparation, and visualization, we used the packages tidyverse version 2.0.0 (Wickham et al., 2019), lubridate version 1.9.2 (Grolemund & Wickham, 2011), the function pirateplot() from the package yarrr version 0.1.5 (Phillips, 2017), and the function random_id() from the package ids version 1.0.1 (FitzJohn, 2017).

For data analysis, we used the following packages and functions: glmer() from the package lme4 version 1.1-32 (Bates et al., 2015) for GLMMs with binomial error distribution, lmer() from the package lmerTest version 3.1-3 (Kuznetsova et al., 2017) for LMMs, brm() from the package brms version 2.20.4 (Bürkner, 2017) for the multinomial mixed model, and vif() from the package car version 3.1-2 (Fox & Weisberg, 2019) to calculate variance inflation factors.

Appendix C

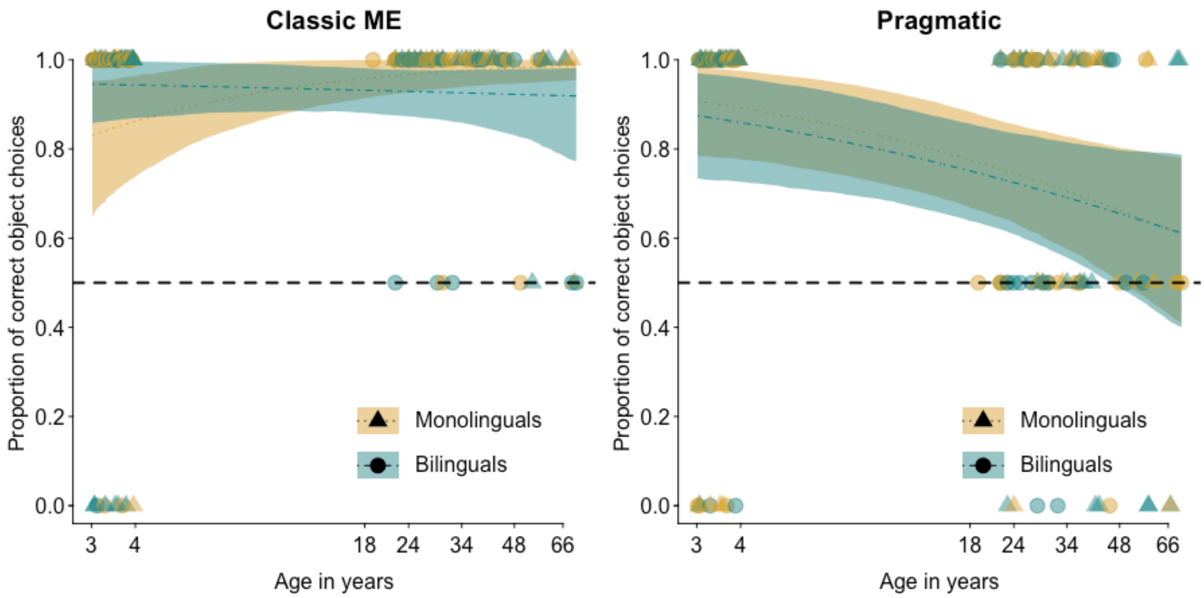
Additional Analyses

Preregistered Analysis of Correct Object Choices in Referent Disambiguation Trials

We preregistered to analyze participants' correct object choices in disambiguation trials in a model that included continuous age as a predictor. To ease the model interpretation, we decided to replace this predictor with the factor age group and reported the results of this model in the main text. In the following, we report the results of the preregistered model which revealed the same pattern of results.

Figure C1

Correct Object Choices in Referent Disambiguation Trials



Note. Grey triangles (monolinguals) and dots (bilinguals) show the proportions of correct object choices per participant, based on trials in which any selection was made (children: $n_{monolingual} = 74$, $n_{bilingual} = 72$; adults: $n_{monolingual} = 167$, $n_{bilingual} = 172$). Dotted (monolinguals) and dash-dotted (bilinguals) lines represent the fitted values; and darker (monolinguals) and lighter (bilinguals) polygons show the 95% confidence intervals, both calculated via bootstrapping with 1000 boots. The values base on the GLMM with binomial error distribution predicting correct choices in disambiguation trials by language background (monolingual/bilingual), condition (Classic ME/Pragmatic), age (continuous, log- and z-transformed), and all of their possible interactions as predictors, the speaker's gaze order (first to target/distractor) as a control variable and random intercepts for participants.

Table C1

Model Predicting Correct Choices in Referent Disambiguation Trials by Language Background, Condition, Continuous Age, Their Interactions, and Gaze Order

	Estimate	*SE*	*p*	95% *CI*
Reference groups: monolinguals, Classic ME condition				
Intercept	3.28	0.54	<.001	2.51, 6.82
Language background	-0.47	0.59	.426	-3.87, 0.86
Condition	-1.83	0.53	<.001	-5.47, -0.92
Z-log-age	0.95	0.38	.013	0.24, 3.38
Gaze order	-0.36	0.27	.190	-0.89, 0.19
Language background x condition	0.33	0.66	.621	-1.20, 3.94
Language background x z-log-age	-1.09	0.53	.038	-4.08, -0.11
Condition x z-log-age	-1.58	0.45	<.001	-4.30, -0.80
Language background x condition x z-log-age	1.21	0.63	.053	-0.08, 4.33
Reference groups: bilinguals, Pragmatic condition				
Intercept	1.31	0.30	<.001	0.81, 2.03
Language background	0.14	0.35	.681	-0.61, 0.96
Condition	1.50	0.41	<.001	0.79, 2.82
Z-log-age	-0.52	0.26	.051	-1.15, -0.03
Gaze order	-0.36	0.27	.190	-0.89, 0.19
Language background x condition	0.33	0.66	.621	-1.20, 3.94
Language background x z-log-age	-0.12	0.37	.747	-0.99, 0.65
Condition x z-log-age	0.37	0.44	.399	-0.98, 1.27
Language background x condition x z-log-age	1.21	0.63	.053	-0.08, 4.35

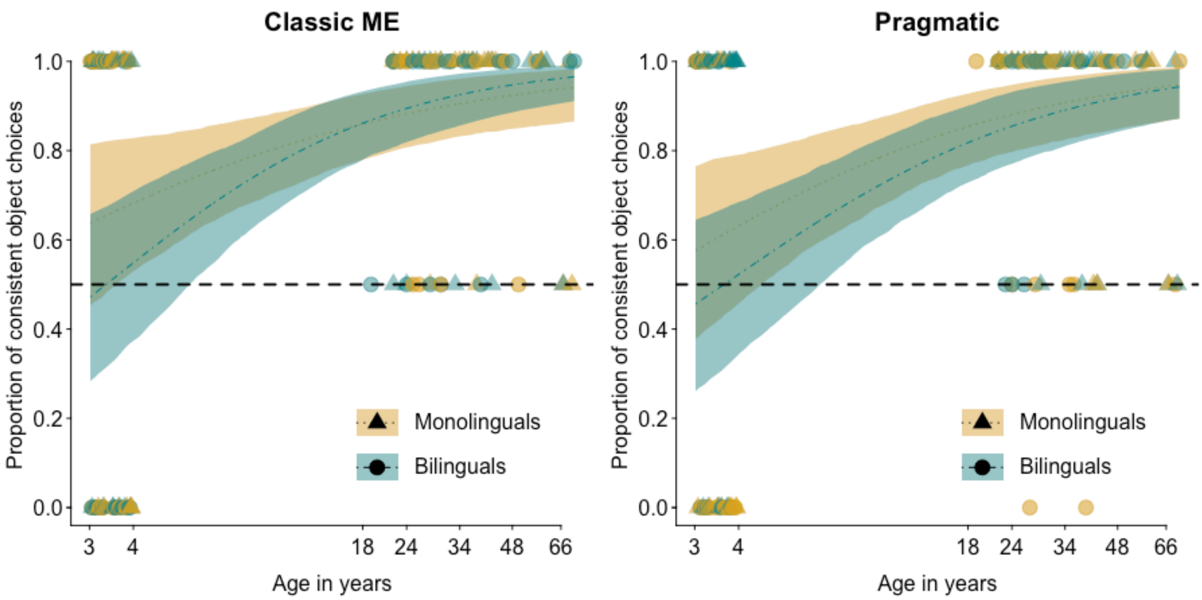
Note. GLMM with binomial error distribution on participants' correct choices in referent disambiguation trials with language background (monolingual/bilingual), condition, continuous age (log- and z-transformed; M = 0, SD = 1) and all of their interactions as predictors, gaze order (first to target/distractor) as control variable and random intercepts for participants (SD = 0.72). $N_{observations}$= 485. N_{groups} = 160. The 95% confidence intervals were obtained via bootstrapping with 1000 boots. The model described the data significantly better than the corresponding null model ($\chi^2(7) = 52.62$, $p < .001$).

Preregistered Analysis of Consistent Object Choices in Retention Trials

We preregistered to analyze participants' consistent object choices in retention trials in a model that included continuous age as a predictor. To ease the model interpretation, we decided to replace this predictor with the factor age group and reported the results of this model in the main text. In the following, we report the results of the preregistered model which revealed the same pattern of results.

Figure C2

Consistent Object Choices in Retention Trials



Note. Grey triangles (monolinguals) and dots (bilinguals) show the proportions of consistent object choices in retention trials per participant, based on trials in which any selection was made (children: $n_{\text{monolingual}} = 73$, $n_{\text{bilingual}} = 70$; adults: $n_{\text{monolingual}} = 166$, $n_{\text{bilingual}} = 172$). Dotted (monolinguals) and dash-dotted (bilinguals) lines represent the fitted values; and darker (monolinguals) and lighter (bilinguals) polygons show the 95% confidence intervals, both calculated via bootstrapping with 1000 boots. The values base on the GLMM with binomial error distribution predicting consistent choices in retention trials by language background (monolingual/bilingual), condition, age (continuous, log- and z-transformed), and all of their possible interactions as predictors, and random intercepts for participants.

Table C2

Model Predicting Consistent Choices in Retention Trials by Language Background, Condition, Continuous Age, Their Interactions, and Gaze Order

	Estimate	*SE*	*p*	95% *CI*
Reference groups: monolinguals, Classic ME condition				
Intercept	1.78	0.31	<.001	1.24, 2.49
Language background	-0.00	0.42	.993	-0.78, 0.87
Condition	-0.05	0.38	.891	-0.79, 0.76
Z-log-age	0.76	0.26	.003	0.27, 1.35
Language background x condition	-0.27	0.54	.625	-1.26, 0.80
Language background x z-log-age	0.42	0.37	.255	-0.32, 1.22
Condition x z-log-age	0.13	0.33	.696	-0.63, 0.83
Language background x condition x z-log-age	-0.29	0.48	.548	-1.25, 0.70
Reference groups: bilinguals, Pragmatic condition				
Intercept	1.46	0.30	<.001	0.93, 2.12
Language background	0.27	0.41	.506	-0.52, 1.12
Condition	0.32	0.38	.408	-0.43, 1.05
Z-log-age	1.02	0.27	<.001	0.54, 1.62
Language background x condition	-0.27	0.54	.625	-1.32, 0.78
Language background x z-log-age	-0.13	0.36	.716	-0.91, 0.58
Condition x z-log-age	0.16	0.35	.647	-0.53, 0.85
Language background x condition x z-log-age	-0.29	0.48	.548	-1.24, 0.68

Note. GLMM with binomial error distribution on participants' consistent choices in retention trials with language background (monolingual/bilingual), condition (Classic ME/Pragmatic), continuous age (log- and z-transformed; $M = 0$, $SD = 1$) and all of their interactions as predictors, and random intercepts for participants ($SD = 0.85$). $N_{observations} = 481$. $N_{groups} = 159$. The 95% confidence intervals were obtained via bootstrapping with 1000 boots. The model described the data significantly better than the corresponding null model ($\chi^2(7) = 51.37$, $p < .001$).

Analysis of Bilinguals in a Broader Sense

In the main analyses, we only included bilingual children in a stricter sense, i.e., children with at least 20% input in their additional language(s). In the following, we report the results of the same analyses including those additional 8 children that we categorized as being bilingual only in a broader sense. These included children who received only 5% (*n*=2) or 10% (*n*=3) input in the additional language(s), children with bilingual input only in their past (*n*=1), and children who were only exposed to the sound of another language via TV/music (*n*=1) or one parent talking in another language to family relatives on the phone, but not to the child/other family members (*n*=1). See Tables C3 and C4 for the results of the full models including all possible interactions of the predictors.

As is the main analysis, we removed non-significant interactions in a stepwise fashion. The final reduced model mirrored the results of the main analysis: Participants' language background did not affect their performance (b = -0.08, SE = 0.29, p = .777), and neither did the speaker's gaze order (b = -0.36, SE = 0.27, p = .182). However, there was a significant interaction of condition and age group (b = -2.59, SE = 0.62, p < .001): That is, children's performance did not differ between conditions (b = 0.30, SE = 0.47, p = .522), while adults performed significantly better in the Classic ME than the Pragmatic condition (b = -2.30, SE = 0.41, p < .001). Similarly, adults performed significantly better than children in the Classic ME condition (b = 1.28, SE = 0.49, p = .009), but significantly worse than children in the Pragmatic condition (b = -1.32, SE = 0.42, p = .002).

Table C3

Analysis of Correct Choices in Referent Disambiguation Trials for Bilinguals in a Broader Sense

	Estimate	*SE*	*p*	95% *CI*
Reference groups = monolinguals, Classic ME condition, children				
Intercept	1.82	0.51	<.001	1.00, 3.17
Language background	0.44	0.65	.504	-1.02, 1.98
Condition	0.43	0.66	.514	-0.95, 2.05
Age group	2.36	0.87	.006	0.87, 11.77
Gaze order	-0.36	0.27	.184	-0.92, 0.17
Language background x condition	-0.26	0.94	.784	-2.69, 1.98
Language background x age group	-1.78	1.07	.095	-11.53, 0.08
Condition x age group	-3.55	1.02	.001	-13.40, -1.60
Language background x condition x age group	1.55	1.30	.233	-1.18, 11.76
Reference groups = bilinguals, Pragmatic condition, adults				
Intercept	1.01	0.32	.002	0.42, 1.70
Language background	0.06	0.39	.886	-0.78, 0.93
Condition	1.83	0.48	<.001	0.96, 2.97
Age group	1.43	0.58	.014	0.36, 3.05
Gaze order	-0.36	0.27	.184	-0.92, 0.17
Language background x condition	1.29	0.90	.151	-0.24, 11.89
Language background x age group	-0.24	0.82	.774	-2.04, 1.65
Condition x age group	-2.01	0.82	.014	-4.06, -0.25
Language background x condition x age group	-1.55	1.30	.233	-12.81, 1.18

Note. GLMM with binomial error distribution on children's and adults' correct choices in referent disambiguation trials with language background, condition, age group and all of their interactions as predictors, speaker's gaze order (first to target/distractor) as control variable (reference group = first to distractor), and random intercepts for participants (SD = 0.76). $N_{observations}$= 501. N_{groups} = 168. The model bases on all mono- and bilinguals, including children categorized as bilingual in a broader sense. The 95% confidence intervals were obtained via bootstrapping with 1000 boots. The model described the data significantly better than the corresponding null model ($\chi^2(7) = 51.32$, $p < .001$).

Table C4

Analysis of Consistent Choices in Retention Trials for Bilinguals in a Broader Sense

	Estimate	*SE*	*p*	95% *CI*
Reference groups = monolinguals, Classic ME condition, children				
Intercept	0.57	0.39	.141	-0.19, 1.38
Language background	-0.58	0.52	.268	-1.66, 0.50
Condition	-0.31	0.51	.547	-1.41, 0.72
Age group	1.80	0.56	.001	0.72, 3.04
Language background x condition	0.09	0.69	.902	-1.40, 1.39
Language background x age group	0.61	0.75	.422	-0.88, 2.32
Condition x age group	0.42	0.73	.565	-0.98, 1.92
Language background x condition x age group	-0.55	1.00	.582	-2.72, 1.28
Reference groups = bilinguals, Pragmatic condition, adults				
Intercept	2.05	0.38	<.001	1.38, 3.00
Language background	0.43	0.53	.415	-0.58, 1.66
Condition	0.35	0.48	.472	-0.60, 1.44
Age group	-2.28	0.53	<.001	-3.51, -1.31
Language background x condition	-0.46	0.72	.518	-2.12, 0.95
Language background x age group	-0.06	0.75	.939	-1.54, 1.48
Condition x age group	-0.12	0.67	.853	-1.52, 1.22
Language background x condition x age group	0.55	1.00	.582	-1.28, 2.72

Note. GLMM with binomial error distribution on children's and adults' correct choices in referent disambiguation trials with language background, condition, age group and all of their interactions as predictors, and random intercepts for participants (SD = 0.80). $N_{\text{observations}}$= 497. N_{groups} = 167. The model bases on all mono- and bilinguals, including children categorized as bilingual in a broader sense. The 95% confidence intervals were obtained via bootstrapping with 1000 boots. The model described the data significantly better than the corresponding null model ($\chi^2(7) = 63.57$, $p < .001$).