

****MANUSCRIPT ACCEPTED BY INFANCY****

Patterns of Mutual Exclusivity and Retention: A Study of Monolingual and Bilingual 2-Year-Olds

Joscelin Rocha-Hidalgo¹, Mary Feller¹, Olivia A. Blanchfield¹, Sarah C. Kucker², and Rachel F. Barr¹

¹Georgetown University, ²Oklahoma State University

Author Note

Correspondence concerning this document should be addressed to Joscelin Rocha-Hidalgo, 3700 P st NW, White Gravenor 306, Washington D.C. 20007. E-mail: jr1679@georgetown.edu

Acknowledgments

We are grateful to all the families who participated in this research and to our research assistants Grace E. Lozano, Madeline Lui, Jenna Ryu, and Olivia Mirek, for their invaluable help. This research was funded by the NSF BCS-1551719. "*From memory flexibility to cognitive flexibility: Examining precursors to bilingual advantages during early childhood*" to Barr.

The authors declare no conflicts of interest with regard to the funding source for this study.

Data and analytic scripts are available at <https://osf.io/syjbz/>. Videos of participants were uploaded to Databrary at <http://doi.org/10.17910/b7.1164>.

Abstract

When children learn their native language, they tend to treat objects as if they only have one label—a principle known as mutual exclusivity. However, bilingual children are faced with a different cognitive challenge—they need to learn to associate two labels with one object. In the present study, we compared bilingual and monolingual 24-month-olds' performance on a challenging and semi-naturalistic forced-choice referent selection task and retention test. Overall, both language groups performed similarly on referent selection but differed on retention. Specifically, while monolingual infants showed some retention, bilingual infants performed at chance and significantly worse than their monolingual peers.

Keywords: Word learning; Bilingualism; Referent Selection; Novelty; Mutual Exclusivity; Disambiguation

Patterns of Mutual Exclusivity and Retention: A Study of Monolingual and Bilingual 2-Year-Olds

By three years of age, children have acquired the major components of their native languages (Adamson, 2018) facilitated by a vocabulary spurt in their second year of life (Samuelson & McMurray, 2017). For example, U.S. English-speaking children go from comprehending and producing a median of 91 words at 18 months to 423 words by 24 months (Frank et al., 2017). Researchers have partly attributed this spurt to children's ability to quickly determine what item a new label refers to (Markman & Wachtel, 1988).

Imagine you are interacting with a toddler, and there are two objects: a duck and a novel object. When asking the toddler to get the *zylow* (or any novel label), the toddler must disambiguate the familiar from the novel items to determine the referent for such a novel name. One way to do so would be to eliminate the familiar toy as a possible referent because it already has a name ("duck"), choosing the novel object instead. Some have attributed this ability to disambiguate novel from known referents to a principle known as mutual exclusivity (ME)—a constraint allowing each item to only have one label (Conboy & Thal, 2006; see Lewis et al. 2020 for a review/meta-analysis). That is because the familiar object had a *duck* label, *zylow* must refer to the novel object. Even though mutual exclusivity is thought to facilitate real-time word learning and vocabulary extension during infancy and toddlerhood (Markman, 1989), few studies test children's ability to retain these novel word-object mappings after some delay (Horst & Samuelson, 2008). Retention is critical as mutual exclusivity (in monolinguals and bilinguals) is only a small part of the word learning process. Furthermore, most of the literature on disambiguation and mutual exclusivity focuses on monolingual populations. Since bilingual infants' word learning environment violates the mutual exclusivity assumption by having two basic-level labels for each object, it is important to understand how bilingual infants solve the disambiguation problem to retain new words and the role of early multilingual language exposure on lexical acquisition.

Monolingualism and Disambiguation

Forced choice selection tasks are the most popular paradigm for measuring disambiguation in the developmental literature (Lewis et al., 2020). The paradigm was introduced by Markman and Watchel

(1988). In their series of studies, children were introduced to a familiar 3D object next to a novel object. Three to four-year-olds chose the unfamiliar object as a referent for a novel label an average of 8 out of 10 times. Children not only had a bias to assign the novel label to a novel object, but they also tended to assign the novel label to a salient property of a familiar object when no other object was present. Markman and Wachtel (1988) interpreted their findings as evidence that children tend to assume that objects have one sole label. If a new label is given, then it must refer to particular aspects of the object. Similar patterns have been consistently reported in various studies with children two years and under (Clark, 1990, Bion et al., 2013; Golinkoff et al., 1992; Markman, 1991; Merriman & Bowman, 1989; Mervis & Bertrand, 1994; Bion, Borovsky, & Fernald, 2013; Horst & Samuelson, 2008; Spiegel & Halberda, 2011; Vlach & Sandhofer, 2012). This pattern has also been reported with young infants when using looking paradigms (for 17- to 18-month-olds see Byers-Heinlein & Werker, 2009, for 17-month-olds see Halberda, 2003; for 15 to 20-month-olds see Markman et al., 2003). Overall, as children develop, their mutual exclusivity abilities increase, and they perform better on disambiguation tasks. Bion et al. (2013), for instance, demonstrated that in a looking paradigm, 18-month-old children struggled with disambiguation and failed to retain labels; 24-month-old children were largely successful at mutual exclusivity but generally failed to retain the labels. By 30-months, children were generally good at both disambiguation and retention trials. Disambiguation and retention performance were also correlated with vocabulary size in this study - as vocabulary increased, so did performance.

Bilingualism and Disambiguation

Studies with a bilingual population have reported conflicting findings on children's ability to disambiguate. Some studies have reported bilinguals performing at the same rate on disambiguation trials as their monolingual peers (see the following for younger than three years: Frank & Poulin-Dubois, 2002; Kalashnikova et al., 2018; Repnik et al., 2021; see the following for older than three years: Davidson & Tell, 2005; Mervis & Bertrand, 1994; Yow et al., 2017) while others have found bilinguals to disambiguate to a lesser extent (Byers-Heinlein, & Werker, 2009, Houston-Price et al., 2010). When bilingual children are provided with a novel label for a familiar object, they are more likely to use that

label as a second label for the item, while their monolingual peers tend to assign such label to a salient property of the item (Kandhadai et al., 2017; see Au & Glusman, 1990 for older children). For example, Kandhadai and colleagues (2017) tested the mutual exclusivity assumption with 18-month-old monolingual and bilingual infants. They argued that differences in mutual exclusivity performance might be due to bilingual infants accepting a second label for a novel object. To test this prediction, infants were familiarized with an image of a dog, accompanied by the word “dog.” Infants were then familiarized with a different-colored (aqua) version of a dog, labeled as a “zabe.” Finally, infants were familiarized with an image of a purple cat, accompanied by the word “cat.” Infants were subsequently tested with an image of an aqua-colored cat and a purple-colored dog, accompanied by the word “zabe.” Monolinguals used the mutual exclusivity hypothesis, looking longer at the aqua-colored cat, but bilinguals instead looking longer at the novel dog. Bilinguals interpreted a novel label for a familiar object as a second label, whereas monolinguals interpreted the novel label as a property of the object: an adjective. That is, the application of the label differed as a function of children's prior word learning experience. Therefore, the authors concluded that there are qualitative differences in bilinguals’ adoption of the mutual exclusivity assumption during early childhood, which changes how bilinguals interpret novel labels (Kandhadai et al., 2017; see also Weatherhead et al., 2021). Such differences in word learning strategies may mean that monolingual and bilingual infants will apply labels differently in other learning contexts.

In another study, Byers-Heinlein and Werker (2009) compared 17- to 18-month-old monolingual, bilingual, and trilinguals, recording their preferential looking time to a novel and a familiar referent presented on a screen. When given a familiar name, all groups looked longer at the familiar object (replicating previous findings with monolingual infants; de Marchena et al., 2011; Golinkoff et al., 1992). However, during the novel target disambiguation trials, monolingual children disambiguated strongly (looked longer at the novel object) while bilinguals only showed marginal disambiguation, and trilinguals showed no evidence of disambiguation. This study was one of the first to demonstrate disambiguation ability differs as a function of early language exposure. Replicating Byers-Heinlein and Werker (2009), Houston-Price et al. (2010) found that 17- and 22-month-old infants in both language groups looked

longer at the familiar objects when the familiar label was used. Contrary to previous findings, the bilingual group did not show evidence of disambiguation abilities (Houston-Price et al., 2010). They suggested the lack of such an ability may be due to bilinguals having smaller English vocabularies than monolingual infants.

In addition to a more language diverse background, another factor that may be affecting the ability to disambiguate is the extensive exposure to lexical overlap (multiple labels referring to one object) that bilingual children are exposed to from an early age (Kandadhai et al., 2017). For example, contrary to their original hypothesis, Kandadhai and colleagues (2017) found that 1.5-year-old bilinguals were more likely than their monolingual peers to accept a second label for a referent. The authors attributed this finding to bilinguals' more extensive experience accepting more than one label for the same object. In contrast, neither Frank and Poulin-Dubois (2002) nor Kalashnikova and colleagues (2019) were able to find any significant differences between monolingual and bilingual two-and-a-half-year-olds on lexical overlap tasks. It seems that in the presence of an ambiguous situation (as in the task from Kandadhai et al., 2017), bilinguals are more likely to accept a second label while monolinguals are not. On the other hand, when the task introduces a non-ambiguous scenario, both groups perform similarly. Additionally, Kalashnikova et al. (2019) found that vocabulary size predicted performance on the lexical overlap task for the monolingual group only. Taken together, these findings suggest that further inquiry into the role of vocabulary in bilingual and monolingual children's word-learning abilities is warranted.

As suggested by Houston-Price and colleagues (2010), it has been proposed that the lexicon structure may play a role in children's ability to use mutual exclusivity and may be driving the different findings reported in the field. For example, among monolinguals, toddlers who had a larger vocabulary size performed better in disambiguation tasks (Bion et al., 2013; Graham, Poulin-Dubois, & Baker, 1998; Lewis et al., 2020). For a bilingual population, some researchers like Byers-Heinlein and Werker (2013) have looked at the relationship between vocabulary size and disambiguation and the role of children's translation equivalents (TEs; the number of words infants know in both languages) may have on this process. The work on the influence of TE's on word learning is mixed (Byers-Heinlein & Werker, 2013;

Frank & Poulin-Dubois, 2002; Houston-Price et al., 2010). In Byers-Heinlein and Werker (2013), children's disambiguation practices varied by the proportion of TEs reported. When children's vocabulary was closer to a one-to-one lexical structure (few TEs), they exhibited a higher use of disambiguation. On the other hand, when children had a lexicon structure closer to a many-to-one model (i.e., vocabulary containing mostly TEs), they failed to practice disambiguation. However, Houston-Price et al. (2010) found no relation between looking behavior in a similar task and a child's translation equivalents (see also Frank & Poulin-Dubois, 2002). Thus, one aim of the current study is to examine the role of TEs on disambiguation. In addition, the present study also examines their impact on retention.

Retention of Novel Labels after Disambiguation

Although disambiguation is a critical step in the word learning process, it is not the same as retention (Kucker et al., 2015). In many cases, disambiguation has been shown to facilitate monolingual children's abilities to retain novel label-object pairings (e.g., Golinkoff et al., 1992; Horst & Samuelson, 2008; Kucker et al., 2018; Bion et al., 2013; Spiegel & Halberda, 2011; Goodman et al., 1998; Mervis & Bertrand, 1994; Spiegel & Halberda, 2001; Wilkinson & Mazzitelli, 2003). Even though some of these studies (Golinkoff et al., 1992; Spiegel & Halberda, 2001; Bion et al., 2013; Spiegel & Halberda, 2011) demonstrate that children can retain novel word-object pairings after a small delay, other ones have found opposite findings when the population is younger than 30 months of age (see Bion et al., 2013; Horst & Samuelson, 2008; Kucker et al., 2018). For example, Horst and Samuelson (2008) demonstrated that 24-month-old monolinguals could not retain words after a five-minute delay even when they successfully disambiguated in a forced-choice task with three objects present (two familiar and one novel). Retention skills of children 30 months or older vary whenever the number of foils is changed (Horst, Scott, & Pollard, 2010; Zosh, Brinster, & Halberda, 2013) or even when 24-month-old children's familiarity with the novel items is manipulated (Kucker & Samuelson, 2012). It is evident that studies with monolingual populations have been able to investigate how solving disambiguation problems are related to children's retention skills. Still, little attention has been given to bilingual populations.

Recent studies with 24-month-old monolingual and bilingual children have found no significant differences in disambiguation performance (Byers-Heinlein et al., 2014; Kalashnikova et al., 2018). In fact, both studies showed evidence that children at this age disambiguate regardless of their language background. To date, only two studies have examined bilingual children's abilities to retain novel labels after acquiring the labels through disambiguation (Kalashnikova et al., 2018; Repnik et al., 2021). Kalashnikova et al. (2018) utilized preferential looking-while-listening paradigms to measure the use of disambiguation and tested *immediate* retention skills. Even though monolinguals could retain the novel words at both 18 and 24 months, bilinguals could only do so at 18 months (Kalashnikova et al., 2018). The authors suggested that by their second year, bilingual children recognize the inefficiency of mutual exclusivity for word learning; therefore, stop using it for language acquisition and use it instead solely for disambiguation. This may explain why bilinguals disambiguated at both 18 and 24 months old but failed to retain the words at 24 months.

In a second, preferential looking-while-listening study, Repnik and colleagues (2021) examined disambiguation in 18- to 30-month old monolingual and multilingual children. In this study, Repnik and colleagues included trials where children were trained on novel items and novel disambiguation trials (i.e., a familiar item vs. a novel item). In the novel training trials, a novel object was presented alone, and the novel label was repeated (novel trained). There were two types of disambiguation trials, familiar disambiguation (e.g., a car vs. a ball) and novel disambiguation trials--a novel item and a familiar item were presented accompanied by a novel label (e.g., *dax*). During retention trials, children were presented with two novel objects. The object label was either a trained label or a disambiguation label. Monolinguals disambiguated at an earlier age than bilinguals. For the trained item, retention was above chance for both monolingual and multilingual children. For the disambiguated item, retention was above chance for monolinguals but at chance for multilinguals. Although disambiguation was related to retention for monolinguals, this was not the case for bilinguals. Vocabulary size was related to bilingual performance. They found that as vocabulary size increased in bilinguals, both disambiguation and retention increased. Repnik and colleagues (2021) concluded that the difference in prior learning

experiences of monolingual and multilingual children shapes the associative network and in-the-moment learning via disambiguation, and it is the combination of these factors that predicts retention. There were, however, some limitations to the study. Repnik and colleagues included both bilingual and trilingual children in the multilingual group. Due to counterbalancing of trial types, some retention trials occurred before the disambiguation trials. Several open empirical questions remain regarding bilingual exposure disambiguation and retention.

Current Study

The current study aims to replicate prior work on mutual exclusivity in bilingual children, expanding past research to include retention trials using a forced-choice paradigm with physical objects. The study was conducted in the home. According to a reanalysis of a dataset used for a meta-analysis on the study of mutual exclusivity, including both monolingual and bilingual samples in the developmental literature (Lewis et al., 2020; visit [OSF-link-here](#) for reanalysis scripts), revealed that out of 146 studies identified, 121 of them (83%) measured children's behaviors in a forced-choice paradigm, and 25 used eye-tracking. From these 121 studies, 62% of the studies reported using real 3D objects as their stimulus (see [metalab.stanford.edu](#)). That is, the use of forced-choice methods with real objects is common in the literature (see also Horst et al., 2010; Wilkinson et al., 2003). In addition, the use of this paradigm for the present study was practical to facilitate participation by children tested in their homes. We examined bilinguals' mutual exclusivity and retention skills utilizing Kucker and colleagues' (2018) forced-choice referent selection task and retention test. Kucker and colleagues tested 18-month-olds on four familiar referent selection trials and four novel referent selection trials. After a five-minute delay, infants were tested on three retention trials. This challenging task requires infants to select between three items. Kucker et al. (2018) found monolingual infants picked novel referents almost universally, demonstrating disambiguation, but showed poor retention.

Importantly, in Kucker et al. (2018), they found that 18-month-old children demonstrated some level of a novelty bias - choosing the novel foil item on familiar referent trials instead of the target. In related work (Kucker et al., 2020), they also found that although 24-month-old children performed at

chance correctly choosing the familiar foil, they did so at levels lower than when no novel item was present. Following this, performance during the novel referent selection trials in the current study will measure disambiguation. In contrast, their performance in the familiar referent selection trials will be used to control for possible novelty bias driving children's disambiguation performance (see Kucker et al., 2018). Finally, three post-delay retention trials will measure children's abilities to remember the novel labels. We hypothesized that monolingual 2-year-old children would disambiguate to a greater extent than their bilingual peers, similar to that found in prior work (Byers-Heinlein & Werker, 2009, 2013; Houston-Price et al., 2010). For the retention trials, we predicted that we would replicate preferential-looking studies' findings with bilinguals (Kalashnikova et al., 2018; Repnik et al., 2021), showing that even if bilinguals do exhibit disambiguation, they will show lower retention scores than their monolingual peers. Additionally, we hypothesized that bilingual children's high performance in the disambiguation and retention trials would be associated with lower translation equivalent scores but larger vocabulary sizes.

Methods

The present study was conducted according to guidelines laid down in the Declaration of Helsinki, with written informed consent obtained from a parent or guardian for each child before any assessment or data collection. All procedures involving human subjects in this study were approved by the Institutional Review Board at Georgetown University.

Participants

A total of 146 24-month-old children (72 females) from families in the DMV area participated from 2016-2019 as part of an ongoing longitudinal study. Children were included in the analyses if they completed at least two trials for each Referent Selection type (Familiar & Novel) and at least 1 Retention trial. Thus, the final sample consisted of 125 children (64 females). Eighty-one (81) children were classified as monolinguals (44 females, $M_{age} = 25.69\text{mo}$, $SD = 0.95$, range = 24.03-27.97mo) and 44 children were classified as bilinguals (20 females, $M_{age} = 25.44\text{mo}$, $SD = 0.67$, range = 24.17-27.97mo). Sixteen additional children were dropped due to being out of the age range (2), fussiness (2), parental interference (2), experimenter error (1), language information missing (1), no significant exposure to

English or Spanish (2), and being exposed to a third language for 10% or more of the time (6). Table 1 provides information for the final sample.

Most of the parents (99.2%) reported their child's race and ethnicity. 76% reported their child to be Caucasian, 20% Mixed, 1.6% Asian/Asian American, 0.8% Native American, and 0.8% African/African American. Caregivers identified 28 children as Latino/a. Participants were primarily from college-educated families, with a mean number of 17.6 years (both parents' average; $SD = 1.36$ years) of education. The families were middle- to high-income, with an average yearly income of \$88,432.30 ($SD = \$28,558.11$) for monolinguals and \$87,772 ($SD = \$33,539.63$) for bilinguals, based on median household income zip code for the family's postcode at the time of participation.

Materials

Bilingualism Measure

Parents were interviewed about who spoke to the child (i.e., mother, father, siblings, nanny, daycare, etc.), in which language, and for how many hours per day, using the Language Exposure Assessment Tool (LEAT, DeAnda et al., 2016). Following previous studies (Barr et al., 2020; Brito & Barr, 2012; Brito et al., 2014, 2015), if a child was exposed to a second language for at least 20% of the time since birth, they were classified as "bilingual." Otherwise, they were labeled "monolingual." See Table 2 for a list of languages for both groups.

MacArthur Communicative Development Inventory Words and Sentences Short Form (MCDI)

We used the Level II form to measure children's productive vocabulary for the dominant and non-dominant languages (Fenson, 2007; see Table 1). From the 100-word list, parents selected which words their child produced or understood within a week of the study. Due to the wide variety of languages, language-specific vocabulary measures were not feasible. Instead, the online survey prompted parents to indicate all the languages the child is exposed to at any time since birth. The survey was created to use the said list of languages and prompted parents with the same English-word list as many times as languages they listed. Each time parents were instructed to select the language-equivalent words their child could produce.

Stimuli

Three items familiar to the child were selected (*duck, ball, cup*) and verified by the parents as being known (Figure 1, left). Labels were replaced as needed (i.e., parent reporting that the child knew the word *ducky* instead of *duck*). These same three items were used throughout the task to ensure that children were indeed familiar with the labels for these specific exemplars and thus could engage in mutual exclusivity (see also Kucker et al., 2018; Horst & Samuelson, 2008 for similar paradigms). Eight unknown, novel items (Figure 1, right) were confirmed as being unfamiliar to the children by the parents. Four novel words were chosen (*roke, cheem, zylow, lorp*) from the Novel Object and Unusual Name (NOUN) Database (Horst & Hout, 2016). Chosen words had no meaning or phonetic similarities to words in the most common languages spoken by families in previous studies (Barr et al., 2020; Brito et al., 2012, 2014, 2015).

Procedure

All sessions took place in the children's homes. The majority of the sessions were performed in English, with only three sessions in Spanish. The child sat across a table from the primary experimenter. The experimenter re-prompted the child if necessary but did not provide feedback. A second experimenter prepared the toys on the tray out of sight of the child. The parents were directed to refrain from interacting with the child. Children were presented with a series of trials and asked to retrieve 1 of 3 items from a tray using either a familiar or novel label (See Figure 2). We used a fixed order presentation of trials (alternating familiar and novel referent selection trials).

Familiar/ Warm-Up Trials (n=4)

The task started with three trials that served as a warm-up to familiarize the child with the task and ensure they understood the instructions. On each trial, the child was presented with the three known items on a tray side-by-side. The primary experimenter asked the child to retrieve one item from the tray (e.g., "Can you get the *cup*?"). Target locations and objects were randomized. Each object was the target once. No praise was given.

Referent Selection Trials (n=8)

All consisted of two familiar items (from the toys used in the Familiar-Warm up trials) and one novel item. In half of the trials, the target item was familiar (e.g., "Can you get the *duck*?"; familiar referent selection). In the other alternating half of the trials, the experimenter asked the child to retrieve a novel item by name (e.g., "Can you get the *zylow*?"; Novel Referent Selection). All target items were asked once, except for one familiar item (duck), which was asked twice.

Break

There was a 5-minute break immediately following referent selection trials. During this time children played with the experimenter or their own toys. No experimental stimuli were present.

Retention Trials (n=3)

Two novel objects that were targets in the novel referent selection trials were presented along with a novel item used as a foil during the familiar referent selection trials.

Preference Trial

Finally, as a manipulation check, a preference trial was performed with novel objects previously used but with no labels. The inclusion of preference trials controlled for the possibility that infants' referent selection responses are due to preferences for specific objects rather than due to disambiguation or retention. The experimenter prompted the child with, "Can you get one?" This trial provided a measurement of children's preferences based on the amount of exposure. It included 1) one novel item that was never labeled and the child has only seen once in the task, 2) a novel item that had been used as a target only once in the task (during Novel RS trials), 3) a novel item that had been used as a target twice (once in Novel RS and once in retention trials). If each item was similarly chosen in a no-label context, this finding would suggest that preference for specific items was not driving the responses.

Results

Coding

All sessions were video recorded for later coding using Datavyu (<http://datavyu.org/>). Coding consisted of timestamping each trial, indicating the pieces laid out on the trays, the label used, final

choice, and any comments associated with the task. We used *R* (Version 3.6.1; R Core Team, 2019) for all analyses.

Inter-coder Reliability

A secondary coder coded 60 videos for reliability purposes using the primary coder's timestamps for each trial. Inter-coder agreement on child's response ($k = 0.94$) was acceptable above 0.70 (Landis & Koch, 1977).

Dependent Variables

To replicate previous work with this task, we divided the total number of times the child chose the target item by the total valid trials per trial type (Invalid = child not answering or refusing to engage with stimuli after at least three attempts from experimenter to attend to the experimental stimuli) to obtain independent scores for familiar (warm-up) trials, familiar referent selection trials, Novel Referent Selection trials, and Retention trials. We refer to this score as "Percent Correct Score."

For the mixed-effect regressions, we used a binary variable denoting whether the child accurately chose the target item (1) or not (0) trial by trial – "Correct Choice."

Preliminary Analyses

Preliminary analyses were conducted to determine whether there was differential attrition between monolinguals and bilinguals, differences in child/family characteristics (age, gender, and SES with parental years of education and income as proxies), differences in English and total vocabulary, and differences in translation equivalents (both total and proportion). We used logistic regression and t-tests to conduct these preliminary analyses.

Attrition Rates

As indicated by preliminary logistic regressions, the attrition rates did not differ as a function of bilingual status (Familiar referent selection trials: $\chi^2(1) = 0.84, p = 0.36$; novel referent selection trials: $\chi^2(1) = 0.004, p = 0.95$; Retention trials: $\chi^2(1) = 0.11, p = 0.74$) and therefore missing data was not a function of bilingual status.

Child and Family Characteristics

Monolingual and bilingual children did not differ in age, female/male ratio, or SES (with parental years of education and income as proxies; see Table 1).

English Vocabulary

Based on the MCDI, monolingual and bilingual children differed significantly on English vocabulary sizes (see Table 1). Vocabulary parental report for English among monolinguals was significantly higher ($M = 66.22$, $SD = 22.64$) compared to bilinguals ($M = 53.44$, $SD = 27.19$), $t(63.76) = 2.11$, $p = 0.02$, $d = 0.51$. Fourteen children did not have a completed MCDI for English and thus were not included in these analyses.

Total Vocabulary

To calculate the total number of words children knew, their raw MCDI scores for their L1 and L2 were summed. Monolingual ($M = 70.77$, $SD = 24.38$) and bilingual ($M = 91.11$, $SD = 48.83$) children had significantly different total vocabulary sizes, $t(44.72) = -2.40$, $p = 0.02$, $d = 0.53$ (see Table 1). Twelve children did not have a completed MCDI for any of their languages and thus were not included in vocabulary analyses.

Translation Equivalents (TEs)

To calculate the total number of translation equivalents, we added the number of MCDI words parents selected for both the child's first and second language. Since some monolinguals had some exposure to multiple languages, translation equivalents were created for them for descriptive purposes ($M = 4.11$, $SD = 9.45$). For the bilingual group, the average number of translation equivalents was 34 words ($SD = 27.83$). To match Byers-Heinlein & Werker (2013)'s approach, the proportion of translation equivalents was included and calculated for both Monolinguals ($M = 5.32$, $SD = 9.79$) and Bilinguals ($M = 30.05$, $SD = 14.40$) by dividing the number of translation equivalents by the total words selected for L1 and L2.

Traditional Analyses

First, we conducted analyses consistent with prior literature using this paradigm. The findings per trial type as a function of bilingual status are presented in Figure 3. For each trial, we calculated whether

the performance was above chance for each group and then compared monolingual and bilingual groups to each other. We ran Levine's test on each analysis to test for equal variance and adjusted degrees of freedom if unequal variances were detected; none were (range p 's 0.10- 0.81).

Familiar (Warm-Up) Trials

Monolinguals performed well on familiar trials, selecting the target 83.98% of the time, well above chance (33.33%), $t(76) = 16.83, p < .00001, d = 1.92$. Bilinguals selected the target item 80.23% of the time, well above chance (33.33%), $t(42) = 10.24, p < .00001, d = 1.56$. Children's accuracy identifying the familiar target was high regardless of language status.

There was no significant difference between monolinguals ($M = 83.98, SD = 26.41$) and bilinguals ($M = 80.23, SD = 30.05$), $t(118) = .71, p = 0.48, d = 0.14$.

Familiar Referent Selection Trials (Familiar RS)

Monolinguals selected the target 46.6% of the time, above chance (33.33%), $t(80) = 3.05, p = 0.003, d = 0.34$. Bilinguals selected the target 43.37% of the time, not above chance (33.33%), $t(43) = 1.75, p = 0.09, d = 0.26$. When comparing monolinguals ($M = 46.6, SD = 39.17$) and bilinguals ($M = 43.37, SD = 38.06$), there was no significant difference, $t(123) = 0.45, p = 0.66, d = 0.08$.

Additionally, we assessed their performance when they failed to choose the target item in order to identify whether their bias was towards the other familiar item or the non-labeled novel item.

Monolinguals selected the novel object 85.38% of the times significantly above chance (50%), $t(60) = 9.99, p < 0.001, d = 1.28$. Bilinguals selected the novel object 86.81% of the times significantly above chance (50%), $t(35) = 7.52, p < 0.001, d = 1.25$. When children did not choose the familiar object in familiar RS trials, there was no significant difference, $t(95) = -0.24, p = 0.81, d = 0.05$, between monolinguals' ($M = 85.38, SD = 27.67$) and bilinguals' ($M = 86.81, SD = 29.37$) rate of selecting the novel toy. Both monolinguals and bilinguals who failed to retrieve the target item did so due to a bias towards the novel item.

Novel Referent Selection trials (Novel RS)

Monolinguals selected the novel object 88.37% of the times significantly above chance (33.33%), $t(80) = 20.83, p < .00001, d = 2.31$. Bilinguals selected the novel object 89.39% of the times significantly above chance (33.33%), $t(43) = 20.06, p < .00001, d = 3.02$. There was no significant difference between monolinguals ($M = 88.37, SD = 23.78$) and bilinguals ($M = 89.39, SD = 18.53$), $t(123) = -0.25, p = 0.81, d = 0.05$.

Retention Trials

Monolinguals selected the novel object 51.23% of the time significantly above chance (33.33%), $t(80) = 5.37, p < .00001, d = 0.6$. Bilinguals selected the novel object 37.88% of the time, not significantly above chance (33.33%), $t(43) = 1.16, p = 0.25, d = 0.17$. Monolinguals performed better in retention trials ($M = 51.23, SD = 30.02$) than bilinguals ($M = 37.88, SD = 26.01$), $t(123) = 2.47, p = 0.01, d = 0.47$.

We also analyzed children's performance when they failed to correctly choose the target novel item in order to test whether the lack of retention was due to a bias towards the never-labeled novel object. Monolinguals selected the non-labeled novel object 33.82% of the time, significantly *below* chance (50%), $t(67) = -3.2, p < 0.002, d = 0.39$. Bilinguals selected the not-labeled novel object 43.65% of the times not significantly different from chance (50%), $t(41) = -1.04, p = 0.30, d = 0.16$. When children failed to choose the right item during retention, monolinguals did not do so due to a bias for the novel object that has never been labeled. They chose the other labeled object at a higher rate. There was no significant difference ($t(108) = -1.23, p = 0.22, d = 0.24$) between monolinguals' ($M = 33.82, SD = 41.63$) and bilinguals' ($M = 43.65, SD = 39.47$) rate of selecting the never-labeled novel toy. This rate of choice was not significantly different between the language groups even though bilinguals were not significantly below chance.

Preference Trial

The preference trial served as a manipulation check to rule out the possibility that children preferred particular novel objects in a non-labeling context. Seventy-four monolingual and 36 bilingual children completed the preference trial. There was no significant difference in the distribution in the

preference of items among monolinguals ($\chi^2(2) = 1.91, p = 0.38$), or bilinguals ($\chi^2(2) = 2.69, p = 0.26$).

This finding indicates that there was not a stimulus effect and that monolinguals and bilinguals did not differ on their preference distribution for the specific items used. Therefore differences in referent selection and retention cannot be attributed to a preference for specific items. The distribution of items chosen in the preference trial was not significantly different between monolinguals and bilinguals ($\chi^2(2) = 1.54, p = 0.46$). Our results indicate that in the absence of a label, search was random.

Vocabulary, Disambiguation, and Retention

Children's ability to disambiguate novel labels (% correct novel RS) was not significantly correlated with either total vocabulary size ($r = 0.15, p = 0.11$), number of translation equivalents ($r = 0.13, p = 0.19$), or proportion of translation equivalents ($r = 0.05, p = 0.65$). In addition, retention of novel labels (% Correct Retention Trials) was not significantly correlated with either Total Vocabulary size ($r = 0.04, p = 0.67$), number of translation equivalents ($r = -0.07, p = 0.48$), or proportion of translation equivalents ($r = -0.14, p = 0.16$).

Stepwise Multiple Regression Framework

Since the previous analyses did not demonstrate a significant difference between Monolinguals and Bilinguals for their Referent Selection trials, the following stepwise multiple regressions focus only on the Retention trials using the Percent Correct Score as the dependent variable.

Binary Language Status

A stepwise multiple regression predicting children's percent correct in the Retention trials was conducted with all predicting variables on the first step (Binary Language Status, Novel Referent Selection Scores, Familiar Referent Selection scores, and Total Vocabulary Size). We used the stepAIC function in the *MASS* R package to select the best model using the Akaike Information Criterion (AIC). Only Bilingual Status entered into the equation at the final step of the analysis ($AIC = 774.60$), $F(1,113) = 4.63, p = .03$. There was a main effect of language status, with bilinguals scoring significantly lower than monolinguals in the retention trials (Table 1 in Supplemental Materials).

Bilingual Exposure as a Continuum

To further qualify the association between second language exposure and retention, a both-direction stepwise multiple regression predicting percent Correct for Retention trials was also conducted with all predicting variables from earlier but using children's L2 exposure instead of the binary bilingual status variable. Only L2 exposure entered into the equation at the final step of the analysis (AIC = 774.18), $F(1,113) = 5.07$, $p = .03$. For every 10% increase in exposure to a second language, children's retention performance decreased by 3.7 % (Table 1 in Supplemental Materials).

Mixed-Effects Logistic Regression Framework

As opposed to the more traditional framework reported above, the mixed-effects regression framework allowed us to model trial-level data from individual infants rather than analyzing averages from their proportion scores. All children who met the criteria contributed to this analysis. Our full fixed effects modeled the main variables of interest: Bilingualism (either binary or continuous), Trial type (for referent selection trials only), Total Vocabulary, and Translational Equivalents (for referent selection trials and retention only) with *participant* and *trial number* as our random effects. However, the best fitting models for each trial type justified by the data were selected comparing AIC values using the *Performance* package (Seedorff et al., 2019).

To assess performance against chance for this trial-by-trial framework, the best fitting models with the trial type and language group of interest set as 0 and the other as 1. The significance of the intercept for each model was then used to assess if accuracy within a condition was greater than chance (33%); because the default intercept assumes .5 as chance, an adjusted intercept was calculated by subtracting $\ln(1/2)$ and dividing by the standard error to get a new Z score.

Familiar/Warm-up Trials

The best-fitting models for the familiar/warm-up trials were:

$$\text{correct_choice} \sim \text{bilingualism}(\text{binary or L2 \%}) + \text{total_vocabulary} + (1|\text{participant}) + (1|\text{trialnumber})$$

There were no significant main effects in either model (see Table 2 in Supplemental Materials). This suggests that the trial performance did not vary by language status (Monolingual vs. Bilingual), L2 % exposure, or total vocabulary size.

Both monolinguals and bilinguals performed significantly above chance (see Table 3).

Referent Selection Trials

The best fitting models for the referent selection trials were:

$\text{correct_choice} \sim \text{bilingualism}(\text{binary or L2 \%}) * \text{trial_type} + \text{total_vocabulary_centered} + \text{child_gender} + \text{parents_education_centered} + (1|\text{participant}) + (1|\text{trialnumber})$

Table 4 shows the coefficient estimates from the best fitting models by bilingualism proxy, and Figure 4 visualizes them. There was a significant main effect of Trial Type (Familiar vs. Novel) in both models. Children performed better during the Novel Referent Selection trials compared to the Familiar Referent Trials. However, there was no significant difference between monolinguals and bilingual children in either trial type. This suggests that the performance on Novel and Familiar referent selection trials did not vary by language status (Monolingual vs. Bilingual) or L2 % exposure. In addition, the centered variable for Total Vocabulary size was significant. In general, children with greater total vocabulary performed better regardless of trial type or language group. Bilingual and monolingual children performed significantly above chance during the Novel Referent Selection trials but only monolinguals performed above chance during the Familiar Referent Selection trials (see Table 3).

We then investigated the role translation equivalents (TEs) would play in these trials by filtering only bilingual children and having *TEs* (Total or Proportion) as our variable of interest (see a similar approach in Houston-Price et al., 2010 and Byers-Heinlein et al., 2013). As with the previous models, there was only a significant effect on *trial type* when Total TEs were included. Children performed better during novel referent selection trials than familiar referent selection trials. Translation equivalents could not predict bilingual children's performance during the familiar and novel referent selection trials even when L2 exposure was included (see Table 3 and Table 4 in Supplemental Materials).

Retention Trials

The best fitting model for the retention trials was:

correct_choice ~ bilingualism(binary or L2 %) + total_vocabulary_centered +
 Familiar_RS_%corr_score_centered + Novel_RS_%corr_score_centered +(1|participant) +
 (1|trialnumber)

Table 5 shows the coefficient estimates from the best fitting models by bilingualism proxy and Figure 5 visualizes them. There was only a significant main effect of bilingualism with monolinguals outperforming their bilingual peers. In the second model, greater exposure to a second language predicted lower retention performance. Children's performance during the Familiar and Novel Referent Selection trials did not predict their performance during the Retention trials. Additionally, only monolingual children performed significantly above chance (see Table 3).

We then investigated the role translation equivalents (TEs) would play in these trials by filtering only bilingual children with *Total TEs* and *Proportion of TEs* variables as our variables of interest. There were no significant main effects even when L2 percent exposure was included. Translation equivalents (proportion or total) did not predict bilingual children's performance during the retention trials (see Table 3 and Table 4 in Supplemental Materials).

Discussion

Our hypotheses were partially supported. Monolingual and bilingual 24-month-old infants did not differ on familiar or novel referent selection but did differ at retention. Specifically, while monolingual infants showed some retention, bilingual infants performed at chance, performing significantly worse than their monolingual peers. These findings extend Kucker and colleagues (2018), who demonstrated good disambiguation performance on novel referent selection but no retention at 18-months. We also find some evidence of a novelty bias on the familiar referent selection trials in both groups - a finding similar to that found in Kucker and colleagues (2018, 2020). This novelty bias suggests that despite knowing an individual word well, children's lexical networks are still fragile, shifting, and developing as their vocabulary grows during the second year of life. Children continued to do well on the novel referent selection trials, however. The current findings of good disambiguation here and greater retention in 24-month-olds suggest that for monolinguals, retention might increase over time following disambiguation.

These results also replicate Kalashnikova and colleagues (2018) results, where 24-month-old monolinguals and bilinguals diverged on their retention patterns following disambiguation in a looking paradigm (see also Repnik et al., 2021). The current findings contribute methodologically to the literature by utilizing the forced-choice paradigm, a more rigorous test of disambiguation with multiple items, and a delayed retention test. Some studies have argued that object manipulation from using 3D items allows the child to form rich representations during learning (Kucker et al., 2020, Pereira et al., 2014; Perry, 2015). It is then easier to attach labels to rich representations during a semi-naturalistic dynamic learning situation, like our task. However, these richer representations were not associated with better retention in bilingual infants.

Despite differences in our methodology, our findings were consistent with those of Kalashnikova et al. (2018), who used a two-item preferential looking task and immediate retention, and those of Repnik and colleagues (2021) who also used a two-item preferential looking task with a 5-minute delay. In addition, the use of a continuous bilingual measure allowed us to explore individual variation in language experience. Bilingual exposure varied greatly in our sample, as demonstrated by the L2%. Our finding that more balanced bilingual infants (as indexed by %L2) showed worse retention, even after a 5-minute break, is consistent with Kalashnikova and colleagues' theory that bilinguals might discount the utility of disambiguation for encoding novel labels. They also hypothesized that children with higher translation equivalents might show poorer retention because of their experience with disambiguation failure. Unlike previous research (Byers-Heinlein & Werker, 2013), we did not find that individual differences in translation equivalents (measured either as total TEs or proportions) were associated with retention. However, a limitation to our protocol was that we used the MCDI short-form, which includes only 100 words reducing the potential variability in the translation equivalent measure. Future research should continue to examine the role that translation equivalents may play in retention.

Although discounting might explain why retention failed, behavioral measures indicated that monolinguals and bilinguals succeeded similarly well in disambiguation trials. Though disambiguation behavior might look the same in monolinguals and bilinguals, the underlying representations and

perceptual processing could potentially differ. That is, selecting a target from an array requires, at a minimum, the exclusion of competitors and focus on the target. Excluding a competitor or foil item to choose a novel requires simply that the other item be minimally familiar (just less so than the target); it does not require deep knowledge or strong semantic representations to do so. For instance, Horst et al. (2011) found that children can disambiguate a “super novel” item from novel ones that were familiarized for less than 30 seconds. Likewise, the lexical trace that is laid down due to a child choosing a target can vary widely. The strength of that lexical trace is dependent upon many factors from the amount of repetition (McMurray et al., 2012), knowledge of competitors (Kucker et al., 2020; Grassman et al., 2020), and even the saliency of the foils (Pomper & Saffran, 2018). Bilingual children here may have failed to retain novel word-referent pairs because the labels for the familiar referents were less well represented in their lexicon due to competition from L2 labels - prior work suggests that the competition present during a referent selection trial predicts the strength of the association formed (McMurray et al., 2012; see also Kucker et al., 2020; Repnik et al., 2021). As such, the word-referent links between the familiar items and their labels were strong enough to facilitate real-time disambiguation but too weak to support robust retention, much like is the case with younger 18-month-olds. Although prior work suggests weaker knowledge can facilitate rich encoding in some cases (Kucker et al., 2020), previous work has studied monolinguals for whom “weak knowledge” is defined as both a weak label and a weak representation of an item (Kucker, & Samuelson, 2012). Bilingual children here may have a weak representation of the English phonological word-form, but still, a rich representation of the referent, making it possible to engage in mutual exclusivity but difficult to draw on the full lexical network in the same way to support retention.

To address these questions, eye tracking could examine whether visual attention allocation differs across monolingual and bilingual infants at different phases of disambiguation and retention. Presumably, if a word is less familiar to an infant, bilingual infants may use more attentional processing to encode, leading to less robust disambiguation and retention. On the other hand, the representational strength of familiar labels could be improved through repetition of the object-word pairings prior to the task. It is also

feasible that bilingual infants may solve the disambiguation trials using visual perceptual discrimination alone - a possibility that such moment-to-moment tests may tap. That is, knowledge for the familiar foils may be so weak infants relied on perceptual novelty to disambiguate instead of vocabulary - selecting the most novel item regardless of label. Likewise, as proposed above, the lexical network on bilingual children may be structured differently from monolingual children. Computational models may help further probe the underlying competitive process supporting these children's word-referent mappings and retention (such as that used by McMurray et al., 2012).

Finally, Repnik and colleagues (2021) suggested that researchers ought to examine the role of other word learning strategies that bilinguals may rely upon. Prior deferred imitation studies have shown that bilingual 24-month-olds outperform monolinguals on memory flexibility tasks. In deferred imitation tasks, during flexibility trials, an experimenter demonstrates target actions on one set of items, but at test, infants are presented with perceptually different but functionally equivalent items. Brito and colleagues (2014) found 24-month-olds monolinguals relied on labels, but bilinguals did not (see Barr et al. 2020 for similar results in 18 month-olds). Taken together, memory flexibility findings suggest that bilinguals may rely more on visual perceptual features and use labels differently than monolinguals. Critically, relying on perceptual features for disambiguation does not typically support word retention in younger 18-month-old children (Kucker et al., 2018) but does facilitate memory flexibility in 18- and 24-months-olds (Barr et al., 2020; Brito et al., 2014). However, because the field has relied on cross-sectional approaches, the developmental trajectory of both mutual exclusivity and memory flexibility is unknown.

In conclusion, our results showing that bilingual toddlers display poorer retention after word-learning in a disambiguation task than monolinguals is consistent with others in the field (Kalashnikova et al., 2018; Repnik et al., 2021). According to dynamic systems theory, differences in early word learning may depend on differences in word learning strategies deployed at the moment, as well as differences in longer-term encoding of the associative semantic network that supports the child's stable lexical access. It may be that differences in the effectiveness of word-learning strategies in combination with the strength of the associations in the underlying lexicon that leads to differences between monolingual and bilingual

retention. Despite these differences, bilingual children rapidly build vocabularies that are structured by overlapping translation equivalents. Future research is still needed to examine alternate heuristics and learning strategies that are effectively deployed by bilingual children, including nonverbal processing differences.

References

- Adamson, L. B. (2018). *Communication development during infancy*. Routledge.
- Au, T. K. F., & Glusman, M. (1990). The principle of mutual exclusivity in word learning: To honor or not to honor?. *Child Development*, 61(5), 1474-1490.
- Barr, R., Rusnak, S. N., Brito, N. H., & Nugent, C. (2020). Actions speak louder than words: Differences in memory flexibility between monolingual and bilingual 18-month-olds. *Developmental science*, 23(2), e12881.
- Bion, R. A., Borovsky, A., & Fernald, A. (2013). Fast mapping, slow learning: Disambiguation of novel word-object mappings in relation to vocabulary learning at 18, 24, and 30 months. *Cognition*, 126(1), 39-53.
- Brito, N., & Barr, R. (2012). Influence of bilingualism on memory generalization during infancy: Bilingualism and memory generalization. *Developmental Science*, 15(6), 812–816.
- Brito, N. H., Grenell, A., & Barr, R. (2014). Specificity of the bilingual advantage for memory: Examining cued recall, generalization, and working memory in monolingual, bilingual, and trilingual toddlers. *Frontiers in Psychology*, 5, 1369. <https://doi.org/10.3389/fpsyg.2014.01369>
- Brito, N. H., Sebastián-Gallés, N., & Barr, R. (2015). Differences in language exposure and its effects on memory flexibility in monolingual, bilingual, and trilingual infants. *Bilingualism: Language and Cognition*, 18(4), 670–682.
- 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.
- 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.
- Byers-Heinlein, K., & Werker, J. F. (2013). Lexicon structure and the disambiguation of novel words: Evidence from bilingual infants. *Cognition*, 128, 407–416.

- Clark, E. V. (1990). On the pragmatics of contrast. *Journal of Child Language*, 17(2), 417-431.
- Conboy, B. T., & Thal, D. J. (2006). Ties between the lexicon and grammar: Cross-sectional and longitudinal studies of bilingual toddlers. *Child Development*, 77(3), 712-735.
- Davidson, D., & Tell, D. (2005). Monolingual and bilingual children's use of mutual exclusivity in the naming of whole objects. *Journal of experimental child psychology*, 92(1), 25-45.
- de Marchena, A., Eigsti, I.-M., Worek, A., Ono, K. E., & Snedeker, J. (2011). Mutual exclusivity in autism spectrum disorders: Testing the pragmatic hypothesis. *Cognition*, 119(1), 96-113.
- DeAnda, S., Bosch, L., Poulin-Dubois, D., Zesiger, P., & Friend, M. (2016). The Language Exposure Assessment Tool: Quantifying language exposure in infants and children. *Journal of Speech Language and Hearing Research*, 59(6), 1346. https://doi.org/10.1044/2016_JSLHR-L-15-0234
- Fenson, L. (2007). *MacArthur-Bates communicative development inventories*. Baltimore, MD: . Paul H. Brookes Publishing Company.
- Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2017). Wordbank: An open repository for developmental vocabulary data. *Journal of Child Language*, 44(2017), 677-694.
- 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.
- Golinkoff, R. M., Hirsh-Pasek, K., Bailey, L. M., & Wenger, N. R. (1992). Children and adults use lexical principles to learn new nouns. *Developmental Psychology*. 28, 99-108.
- Goodman, J. C., McDonough, L., & Brown, N. B. (1998). The Role of semantic context and memory in the acquisition of novel nouns. *Child Development*, 69(5), 1330-1344.
- Graham, S. A., Poulin-Dubois, D., & Baker, R. K. (1998). Infants' disambiguation of novel object words. *First Language*, 18(53), 149-164.
- 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.
- Halberda, J. (2003). The development of a word-learning strategy. *Cognition*, 87(1), B23-B34.

- 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.
- Horst, J. S., & Samuelson, L. K. (2008). Fast mapping but poor retention by 24-month-old infants. *Infancy*, 13(2), 128–157.
- Horst, J. S., Samuelson, L. K., Kucker, S. C., & McMurray, B. (2011). What's new? Children prefer novelty in referent selection. *Cognition*, 118(2), 234–244.
- Horst, J. S., Scott, E. J., & Pollard, J. A. (2010). The role of competition in word learning via referent selection. *Developmental Science*, 13(5), 706–713.
- Houston-Price, C., Caloghiris, Z., & Raviglione, E. (2010). Language experience shapes the development of the mutual exclusivity bias. *Infancy*, 15(2), 125–150.
- 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.
- Kalashnikova, M., Oliveri, A., & Mattock, K. (2019). Acceptance of lexical overlap by monolingual and bilingual toddlers. *International Journal of Bilingualism*, 23(6), 1517–1530.
- Kandhadai, P., Hall, D. G., & Werker, J. F. (2017). Second label learning in bilingual and monolingual infants. *Developmental science*, 20(1), e12429.
- Kucker, S. C., McMurray, B., & Samuelson, L. K. (2015). Slowing down fast mapping: Redefining the dynamics of word learning. *Child Development Perspectives*, 9(2), 74–78.
<https://doi.org/10.1111/cdep.12110>
- Kucker, S. C., McMurray, B., & Samuelson, L. K. (2018). Too much of a good thing: how novelty biases and vocabulary influence known and novel Referent selection in 18-month-old children and associative learning models. *Cognitive Science*, 42, 463–493. <https://doi.org/10.1111/cogs.12610>
- Kucker, S. C., McMurray, B., & Samuelson, L. K. (2020). Sometimes it is better to know less: How known words influence referent selection and retention in 18- to 24-month-old children. *Journal of Experimental Child Psychology*, 189, 104705.

- Kucker, S. C., & Samuelson, L. K. (2012). The first slow step: Differential effects of object and word-form familiarization on retention of fast-mapped words. *Infancy*, 17(3), 295-323.
- Landis, J. R., & Koch, G. G. (1977). An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics*, 363-374.
- 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.
- Markman, E. M. (1991). 3. The whole-object, taxonomic, and mutual exclusivity assumptions as initial constraints. *Perspectives on language and thought: Interrelations in development*, 72.
- 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.
- 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.
- 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.
- Mervis, C. B., & Bertrand, J. (1994). Acquisition of the novel name-nameless category (N3C) Principle. *Child Development*, 65(6), 1646–1662.
- Pereira, A. F., Smith, L. B., & Yu, C. (2014). A bottom-up view of toddler word learning. *Psychonomic bulletin & review*, 21(1), 178-185.
- Perry, L. K. (2015). To have and to hold: looking vs. touching in the study of categorization. *Frontiers in psychology*, 6, 178.
- Pomper, R., & Saffran, J. R. (2019). Familiar object salience affects novel word learning. *Child development*, 90(2), e246-e262.
- R Core Team. (2019). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <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.
- Samuelson, L. L., & McMurray, B. (2017). What does it take to learn a word? *Wiley Interdisciplinary Reviews. Cognitive Science*, 8(1–2).
- Seedorff, M., Oleson, J., & McMurray, B. (2019, January 25). Maybe maximal: Good enough mixed models optimize power while controlling Type I error.
- Spiegel, C., & Halberda, J. (2011). Rapid fast-mapping abilities in 2-year-olds. *Journal of Experimental Child Psychology*, 109(1), 132–140.
- Vlach, H. A., & Sandhofer, C. M. (2012). Fast mapping across time: Memory processes support children's retention of learned words. *Frontiers in psychology*, 3, 46.
- Weatherhead, D., Kandhadai, P., Hall, D. G., & Werker, J. F. (2021). Putting Mutual Exclusivity in Context: Speaker Race Influences Monolingual and Bilingual Infants' Word-Learning Assumptions. *Child Development*.
- Wilkinson, K. M., & Mazzitelli, K. (2003). The effect of 'missing' information on children's retention of fast-mapped labels. *Journal of Child Language*, 30(1), 47-73.
- Wilkinson, K. M., Ross, E., & Diamond, A. (2003). Fast mapping of multiple words: Insights into when "the information provided" does and does not equal "the information perceived." *Journal of Applied Developmental Psychology*, 24(6), 739–762.
- Yow, W. Q., Li, X., Lam, S., Gliga, T., Chong, Y. S., Kwek, K., & Broekman, B. F. (2017). A bilingual advantage in 54-month-olds' use of referential cues in fast mapping. *Developmental Science*, 20(1), e12482.
- Zosh, J. M., Brinster, M., & Halberda, J. (2013). Optimal contrast: Competition between two referents improves word learning. *Applied Developmental Science*, 17(1), 20-28.

Table 1*Participants' Demographics*

Variable	Monolinguals				Bilinguals				P	Cohen's d
	M	SD	Min	Max	M	SD	Min	Max		
Age _{Mo}	25.44	0.67	24.17	27.97	25.69	0.95	24.03	27.97	0.13	0.30
MCDI _{Eng} (raw)	66.22	22.64	10.00	100	53.44	27.19	9	100	0.02*	0.51
MCDI _{Total} (raw)	70.77	24.38	15	155	91.11	48.83	17	187	0.02*	0.53
Total TEs	4.11	9.45	0	63	34	27.83	0	89	0.00**	1.44
Proportions of TEs	5.32	9.79	0	40.65	30.05	14.40	0	47.59	0.00**	2.01
Income [◇] _{USD}	88,432.30	28,558.11	33,135	161,886	87,772.7	33,539.63	33,135	158,766	0.91	0.02
L1 _%	95.16	7.11	75.00	100	63.73	9.78	47.70	79	0.00**	3.68
L2 _%	4.42	6.23	0	19.90	35.39	9.16	20	49	0.00**	3.96
Education Years Avg both	17.62	1.37	14	20	17.57	1.35	14.00	20.00	0.85	0.04

Note. ◇ Calculations from median-zip code-based annual household income. $p < .05^*$, $p < .01^{**}$. TEs = Translation Equivalents

Table 2*Participants' Language Exposure*

	Monolinguals	Bilinguals
L1	English (n = 80) Spanish (n = 1)	English (n = 28) Spanish (n = 9) French (n = 3) Farsi, German, Portuguese, & Russian (n = 1 each)
L2	No L2 (n = 35) Spanish (n = 25) French (n = 5) Urdu (n = 3) Mandarin (n = 2) English, Ga, German, Hindi, Hungarian, Korean, Polish, Portuguese, Sanskrit, Sinhala, Tagalog, & Vietnamese (n = 1 each)	English (n = 17) Spanish (n = 13) French (n = 3) German (n = 2) Afrikaans, Arabic, Cantonese, Farsi, Hindi, Korean, Mandarin, Portuguese, & Vietnamese (n = 1 each)

Note. Languages that infants were exposed to as a function of language status.

Table 3*Performance against Chance for Trial-by-Trial Data for Each Trial Type*

Trial Type	<i>Estimate</i>	<i>SE</i>	<i>Z</i>	<i>p</i>
Familiar (Warm-Up Trials)				
Monolingual	1.91	0.34	5.59	<.001
Bilingual	1.96	0.42	4.70	<.001
Familiar Referent Selection				
Monolingual	-0.15	0.26	-0.56	0.04
Bilingual	-0.26	0.36	-0.74	0.22
Novel Referent Selection				
Monolingual	2.73	0.38	7.25	<.001
Bilingual	2.68	0.47	5.66	<.001
Retention				
Monolingual	0.02	0.32	0.05	0.03
Bilingual	-0.59	0.36	-1.62	0.76

Note. All models included a random effect of subject and trial number. Only the intercept (which was adjusted for chance at 33%) was used to assess performance in each model.

Table 4

Mixed Effects Models Predicting Performance in Referent Selection Trials with Binary Language Status or L2 % Exposure as Coefficients

Model & Predictors	Estimate	SE	Z	p	AIC
<i>Binary Language Status (logLik= -380.8, deviance = 761.6, df.resid = 813)</i>					779.6
Intercept	-0.27	0.32	-0.82	0.41	
Language Status	-0.09	0.41	-0.23	0.82	
Trial Type (Familiar vs. Novel)	2.99	0.35	8.47	<.001	
Total Vocabulary	0.01	0.01	2.08	0.04	
Gender	0.22	0.35	0.62	0.53	
Parental Education (years)	0.23	0.13	1.76	0.08	
Trial Type x Language Status	0.04	0.49	0.08	0.93	
<i>Continuous L2 % (logLik= -380.5, deviance = 760.9, df.resid = 813)</i>					778.9
Intercept	-0.23	0.34	-0.60	0.55	
L2 % Exposure	-0.01	0.01	-0.52	0.60	
Trial Type (Familiar vs. Novel)	2.84	0.37	7.76	<.001	
Total Vocabulary	0.01	0.01	2.09	0.04	
Gender	0.22	0.35	0.61	0.54	
Parental Education (years)	0.23	0.13	1.73	0.08	
Trial Type x Language Status	0.01	0.01	0.83	0.41	

Note. The following variables were mean-centered: Total vocabulary, Parental Education (years). Random effects variables for the final model with binary variable: participant (variance = 2.11; *SD* = 1.45) and trial number (variance = 0.07; *SD* = 0.26). Random effects variables for the final model with L2% variable: participant (variance = 2.11; *SD* = 1.45) and trial number (variance = 0.07; *SD* = 0.26).

Table 5

Mixed Effects Models Predicting Trial-by-Trial Performance in Retention Trials with Binary Language Status or L2 % Exposure as Coefficients

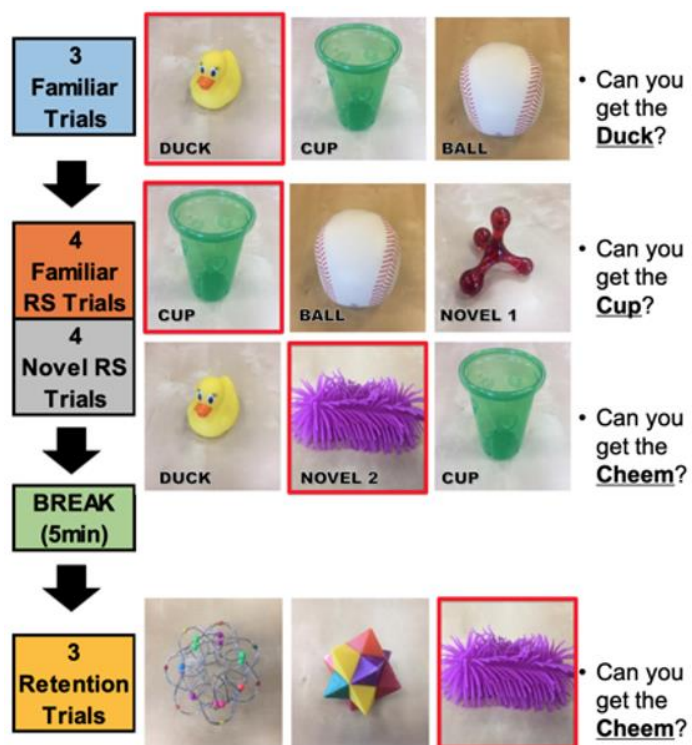
Model & Predictors	Estimate	SE	Z	p	AIC
A. Full models					
<i>Binary Language Status (logLik= -204.3, deviance = 408.7, df.resid = 300)</i>					426.7
Intercept	0.04	0.35	0.11	0.91	
Language Status	-0.60	0.27	-2.23	0.03	
Total Vocabulary	0.003	0.003	0.91	0.36	
Gender	-0.04	0.25	-0.16	0.87	
Parental Education (years)	0.001	0.01	-0.16	0.99	
Familiar RS Percent Score	-0.001	0.003	-0.17	0.87	
Novel RS Percent Score	-0.00	0.01	-0.16	0.87	
<i>Continuous L2 % (logLik= -204.2, deviance = 408.5, df.resid = 300)</i>					426.5
Intercept	0.08	0.36	0.23	0.82	
L2 % Exposure	-0.02	0.01	-2.27	0.02	
Total Vocabulary	0.003	0.003	0.93	0.35	
Gender	-0.01	0.25	-0.05	0.96	
Parental Education (years)	-0.02	0.10	-0.19	0.85	
Familiar RS Percent Score	-0.001	0.003	-0.24	0.81	
Novel RS Percent Score	-0.00	0.01	-0.06	0.95	
B. Final Models					
<i>Binary Language Status (logLik= -204.4, deviance = 408.7, df.resid = 302)</i>					422.7
Intercept	0.02	0.32	0.05	0.96	
Language Status	-0.60	0.27	-2.22	0.03	
Total Vocabulary	0.003	0.003	0.90	0.37	
Familiar RS Percent Score	-0.001	0.003	-0.20	0.84	
Novel RS Percent Score	-0.001	0.01	-0.15	0.88	
<i>Continuous L2 (logLik= -204.3, deviance = 408.5, df.resid = 302)</i>					422.5

Intercept	0.07	0.33	0.22	0.82
L2 % Exposure	-0.02	0.01	-2.26	0.02*
Total Vocabulary	0.003	0.004	0.93	0.36
Familiar RS Percent Score	-0.001	0.003	-0.28	0.78
Novel RS Percent Score	-0.00	0.01	-0.07	0.94

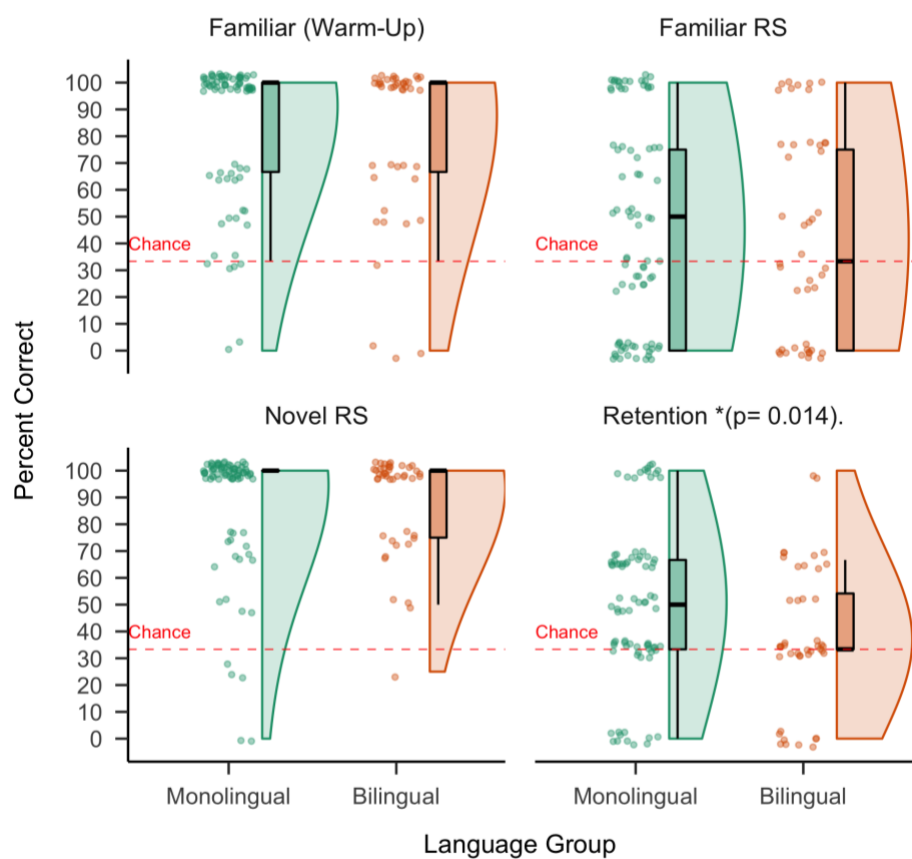
Note. The following variables were mean-centered: Total vocabulary, Parental Education (years), Familiar RS Percent Score, Novel RS Percent Score. Random effects variables for the final model with binary variable: participant (variance = 0.00; *SD* = 0.00) and trial number (variance = 0.25; *SD* = 0.50). Random effects variables for the final model with L2% variable: participant (variance = 0.00; *SD* = 0.00) and trial number (variance = 0.25; *SD* = 0.50).

Figure 1*Experimental Stimuli*

Note: Familiar (left) and novel (right) objects utilized.

Figure 2*Referent Selection Task Procedure*

Note. Procedure used for the Referent Selection Task.

Figure 3*Percent Correct by Language Group across Trials*

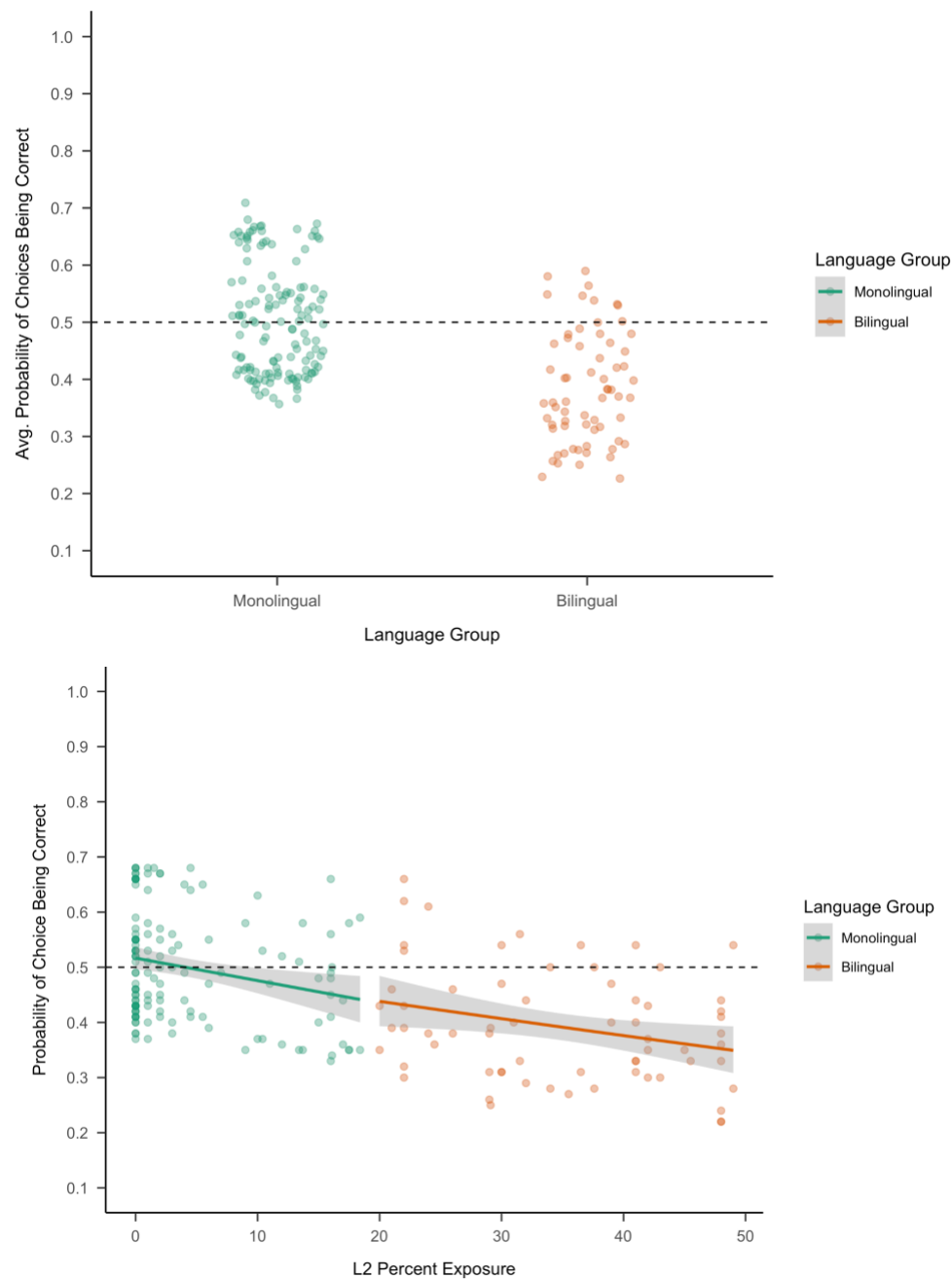
Note. Percent correct by language groups across trials. The red dotted line represents chance.

Figure 4*Visualization of the Logistic Mixed Effect Model for the Referent Selection Trials*

Note. The logistic regression model predicting a child's average probability of a choice being correct during the Referent Selection trials. Thus, chance is represented by a dotted line at 50%.

Figure 5

Visualization of the Logistic Mixed Effect Models for the Retention Trials using both the binary (top) and continuous (bottom) language variables



Note. The logistic regression model predicting a child's average probability of a choice being correct vs. incorrect during the Referent Selection trials. Thus, chance is represented by a dotted line at 50%.

Supplemental Materials

Table 1

Full and Final Models after Both-Direction Stepwise Multiple Linear Regressions Predicting Performance in Retention Trials with Binary Language Status or L2 % Exposure as Coefficients

Model & Predictors	Estimate	SE	t	p	AIC
A. Full models					
<i>Binary Language Status</i>					782.57
Intercept	44.78	37.50	1.19	0.24	
Language Status	-14.29	6.07	-2.35	0.02	
Familiar RS performance	-0.06	0.07	-0.81	0.42	
Novel RS performance	-0.03	0.12	-0.24	0.81	
Total Vocabulary	0.10	0.08	1.20	0.23	
Gender	-1.00	5.72	-0.18	0.86	
Parental Education (years)	0.26	2.14	0.12	0.90	
<i>Continuous L2 %</i>					781.83
Intercept	51.51	37.58	1.37	0.17	
L2 % Exposure	-0.44	0.18	-2.50	0.01	
Familiar RS performance	-0.07	0.07	-0.90	0.37	
Novel RS performance	-0.01	0.12	-0.11	0.92	
Total Vocabulary	0.11	0.08	1.30	0.20	
Gender	-0.28	5.68	-0.05	0.96	
Parental Education (years)	-0.15	2.14	-0.07	0.94	
B. Final Models					
<i>Binary Language Status</i>					774.60
Intercept	50.64	3.26	15.55	<.001	
Language Status	-12.35	5.74	-2.15	0.03	
<i>Continuous L2</i>					774.18
Intercept	51.92	3.55	14.62	<.001	
L2 % Exposure	-0.37	0.17	-2.25	0.03	

Table 2

Mixed Effects Models Predicting Performance in Familiar (Warm Up) Trials with Binary Language Status or L2 % Exposure as Coefficients

Model & Predictors	<i>Estimate</i>	<i>SE</i>	<i>Z</i>	<i>p</i>	<i>AIC</i>
A. Full models					
<i>Binary Language Status (logLik= -118.0, deviance = 237.5, df.resid = 272)</i>					251.5
Intercept	1.80	0.40	4.46	<.001	
Language Status	0.07	0.43	0.17	0.87	
Total Vocabulary	0.001	0.01	0.21	0.84	
Gender	0.18	0.40	0.45	0.66	
Parental Education (years)	-0.09	0.15	-0.62	0.54	
<i>Continuous L2 % (logLik= -118.4, deviance = 236.8, df.resid = 272)</i>					250.8
Intercept	2.01	0.44	4.62	<.001	
L2 % Exposure	-0.01	0.01	-0.88	0.38	
Total Vocabulary	0.003	0.01	0.52	0.60	
Gender	0.15	0.40	0.37	0.71	
Parental Education (years)	-0.09	0.15	-0.64	0.52	
B. Final Models					
<i>Binary Language Status (logLik= -119.0, deviance = 238, df.resid = 274)</i>					248.0
Intercept	1.91	0.34	5.59	<.001	
Language Status	0.05	0.43	0.11	0.91	
Total Vocabulary	0.001	0.01	0.23	0.82	
<i>Continuous L2 (logLik= -118.6, deviance = 237.2, df.resid = 274)</i>					247.2
Intercept	2.09	0.38	5.54	<.001	
L2 % Exposure	-0.01	0.01	-0.88	0.38	
Total Vocabulary	-0.003	0.01	0.53	0.60	

Note. The following variables were mean-centered: Total vocabulary, Parental Education (years). Random effects variables for the final model with binary variable: participant (variance = 0.65; *SD* = 0.81) and trial number (variance = 0.04; *SD* = 0.21). Random effects variables for the final model with L2% variable: participant (variance = 0.67; *SD* = 0.82) and trial number (variance = 0.04; *SD* = 0.20).

Table 3

Mixed Effects Models Predicting Performance in Referent Selection and Retention trials using Total Translation Equivalents (TEs)

Trial Type	Estimate	SE	Z	p	AIC
A. Referent Selection Trials					
<i>Model 1 (logLik= -113.2, deviance = 226.3, df.resid = 231)</i>					238.3
Intercept	-0.44	0.50	-0.89	0.37	
Total Translation Equivalents	0.01	0.01	0.99	0.32	
Trial Type	1.94	0.68	2.85	0.004	
Total TEs x Trial Type	0.03	0.02	1.49	0.14	
<i>Model 2 (logLik= -112.4, deviance = 224.7, df.resid = 230)</i>					238.7
Intercept	0.66	0.99	0.67	0.50	
Total Translation Equivalents	0.01	0.01	1.27	0.21	
Trial type	1.92	0.68	2.80	0.005	
L2 Percent Exposure	-0.03	0.03	-1.26	0.21	
Total TEs x Trial Type	0.03	0.02	1.51	0.13	
B. Retention Trials					
<i>Model 1 (logLik= -53.6, deviance = 107.2, df.resid = 85)</i>					115.2
Intercept	-0.78	0.69	-1.13	0.26	
Total Translation Equivalents	0.01	0.01	0.60	0.55	
<i>Model 2 (logLik= -53.2, deviance = 106.4, df.resid = 84)</i>					116.4
Intercept	-0.01	1.08	-0.01	1.00	
Total Translation Equivalents	0.01	0.01	0.83	0.41	
L2 Percent Exposure	-0.02	0.03	0.93	0.35	

Note. Random effects variables for the models 1 and 2 in the Referent Selection Trial: participant (variance = 1.03, 0.93; *SD* = 1.02, 0.97) and trial number (variance = 0.16, 0.16; *SD* = 0.40, 0.40). Random effects variables for the models 1 and 2 for the Retention trials: participant (variance = 0.00, 0.00; *SD* = 0.00, 0.00) and trial number (variance = 0.93, 0.97; *SD* = 0.97, 0.98).

Table 4

Mixed Effects Models Predicting Performance in Referent Selection and Retention trials using Proportion of Translation Equivalents (TEs)

Trial Type	Estimate	SE	Z	p	AIC
A. Referent Selection Trials					
<i>Model 1 (logLik= -113.1, deviance = 226.2, df.resid = 231)</i>					238.2
Intercept	-0.68	0.71	-0.96	0.34	
Proportion of Translation Equivalents	0.02	0.02	0.95	0.34	
Trial Type	1.51	0.88	1.71	0.09	
Proportion of TEs x Trial Type	0.03	0.02	1.49	0.12	
<i>Model 2 (logLik= -112.2, deviance = 224.5, df.resid = 230)</i>					238.5
Intercept	0.38	1.06	0.36	0.72	
Proportion of Translation Equivalents	0.03	0.02	1.28	0.20	
Trial type	1.53	0.88	1.74	0.08	
L2 Percent Exposure	-0.04	0.03	-1.31	0.19	
Proportion of TEs x Trial Type	0.04	0.03	1.53	0.13	
B. Retention Trials					
<i>Model 1 (logLik= -53.8, deviance = 107.6, df.resid = 85)</i>					115.6
Intercept	-0.63	0.84	-0.75	0.45	
Proportion of Translation Equivalents	0.00	0.02	0.07	0.94	
<i>Model 2 (logLik= -53.5, deviance = 106.9, df.resid = 84)</i>					116.9
Intercept	-0.04	1.12	-0.04	0.97	
Proportion of Translation Equivalents	0.01	0.02	0.33	0.74	
L2 Percent Exposure	-0.02	0.03	-0.80	0.43	

Note. Random effects variables for the models 1 and 2 in the Referent Selection Trial: participant (variance = 1.11, 1.00; *SD* = 1.00, 0.97) and trial number (variance = 0.16, 0.16; *SD* = 0.40, 0.40). Random effects variables for the models 1 and 2 for the Retention trials: participant (variance = 0.00, 0.00; *SD* = 0.00, 0.00) and trial number (variance = 0.94, 0.96; *SD* = 0.97, 0.98).