Look at that: Spatial deixis reveals experience-related differences in prediction

Tracy Reuter, **Mia Sullivan**, **Casey Lew-Williams**
Princeton University, Department of Psychology

Abstract

Prediction-based theories posit that interlocutors use prediction to process language efficiently and to coordinate dialogue. The present study evaluated whether listeners can use spatial deixis (i.e., *this, that, these,* and *those*) to predict the plurality and proximity of a speaker's upcoming referent. In two eye-tracking experiments with varying referential complexity (*N* = 168), native English-speaking adults, native English-learning 5-year-olds, and non-native English-learning adults viewed images while listening to sentences with or without informative deictic determiners, e.g., *Look at the/this/that/these/those wonderful cookie(s).* Results showed that all groups successfully exploited plurality information. However, they varied in using deixis to anticipate the proximity of the referent; specifically, L1 adults showed more robust prediction than L2 adults, and L1 children did not show evidence of prediction. By evaluating listeners with varied language experiences, this investigation helps refine proposed mechanisms of prediction, and suggests that linguistic experience is key to the development of such mechanisms.

Keywords

language processing; deixis; prediction; simulation; association; covert imitation

A number of recent theories claim that prediction can support language processing and learning (Chang, Dell, & Bock, 2006; Christiansen & Chater, 2016; Pickering & Gambi, 2018; Pickering & Garrod, 2007, 2013). For example, Pickering and Garrod (2007; 2013) propose that comprehenders predict upcoming speech in order to coordinate dialogue. Supporting this view, a number of studies demonstrate that both adults and children can generate predictions during language processing (for review see Kutas, DeLong, & Smith, 2011). Despite the rapid pace of spoken language, listeners can use a variety of linguistic and nonlinguistic indications to predict upcoming information in speech. For example, in an eye-tracking experiment, Lukyanenko and Fisher (2016) found that two- and three-year-old children used number markings (*is* and *are*) to predict singular or plural referents in sentences such as *"Where is the good apple?"* and *"Where are the good cookies?"* This finding, among many others, supports the central claim in a number of contemporary psycholinguistic theories that prediction occurs during language processing.

Correspondence should be addressed to: Tracy Reuter, Department of Psychology, Princeton University, Princeton, NJ, 08544. treuter@alumni.princeton.edu.

However, language scientists and developmental scientists have continued to debate a range of issues, such as the extent to which prediction supports everyday language processing (Huettig & Mani, 2016; Kutas, DeLong, & Smith, 2011), and whether prediction supports language acquisition (Phillips & Ehrenhofer, 2015; Rabagliati, Gambi, & Pickering, 2016). A particularly central issue is understanding how prediction occurs at the intersection of language comprehension and production. Pickering and Garrod (2013) proposed that comprehenders can generate predictions via two routes: *association* or *simulation*. The association route relies on comprehension mechanisms, whereas the simulation route relies on production mechanisms. Importantly, for prediction to occur via simulation, the listener "must be able to represent what the speaker would say, not what he himself would say, and to do this, he needs to take into account the context" (Pickering & Garrod, 2013, p. 341). That is, accurate prediction via simulation requires some consideration for the speaker's perspective, such as their visual viewpoint within the referential context, their state of mind, their age, or their knowledge of the conversational topic. Comprehenders can either estimate what their conversational partner would be likely to produce within a particular conversational context (i.e., simulation), or they can generate predictions based strictly on their own perspective (i.e., association).

While Pickering and Garrod (2007; 2013) largely focus on the state of prediction in adulthood, they do speculate as to the development of predictive mechanisms, suggesting that children, ostensibly due to overall differences in language proficiency as compared to native-speaking adults, may not predict via simulation and may instead rely on the association route. Indeed, there are a number of reasons to expect that simulation of a speaker's upcoming utterances may be challenging for children. First, prior findings suggest that while five-year-old children can take a speaker's perspective into account, they still make egocentric errors regarding what is or is not common ground among interlocutors (Clark, 1992; Nilsen & Graham, 2009). If taking the speaker's perspective during language processing is generally challenging for children, then they may have difficulty simulating speakers' perspectives and upcoming words, either deterministically or under certain circumstances (Pickering & Garrod, 2013; Pickering & Gambi, 2018).

Another reason to think that children may not predict via simulation is that they often have difficulty using contextual factors to guide their real-time language processing, such as the number of available referents in a shared visual context, the relations between agents and objects, or the order in which speakers mention referents. Whereas adults are adept in using surrounding visual and discourse context to rapidly resolve linguistic ambiguities, children have difficulty doing so for ambiguous syntax (Hurewitz et al., 2000; Snedeker & Trueswell, 2004; Trueswell, Sekerina, Hill, & Logrip, 1999) and ambiguous pronouns (Arnold, Brown-Schmidt, & Trueswell, 2007). For example, Arnold and colleagues (2007) found that adults combined grammatical gender cues with pragmatic reliance on order-of-mention cues to resolve ambiguous pronouns and accurately identify a speaker's intended referent, but 3- to 5-year-old children failed to take order-of-mention cues into account. Children's relative difficulty in integrating surrounding contextual information could derail rapid, accurate simulation of speakers' upcoming productions.

On the other hand, there are reasons to suspect that children may be capable of generating predictions via simulation. A limited number of prior findings suggest that children can, under some circumstances, incorporate information about the speaker and the surrounding visual and linguistic context when generating predictions. For example, children can generate accurate predictions on the basis of a speaker's disfluencies (Kidd, White, & Aslin, 2011), as well as the speaker's identity (Borovsky & Creel, 2014). Thus, while Pickering and Garrod (2013) speculate that children may be limited to the association route, some results suggest that children (like adults) could use both association and simulation routes for prediction. Additionally, as a general point, children have shown the ability to predict in dozens of prior studies on language processing, suggesting broad prowess in using diverse cues to rapidly interpret referential contexts (Borovsky, Elman, & Fernald, 2012; Borovsky & Creel, 2014; Fernald, Zangl, Portillo, & Marchman, 2008; Havron, de Carvalho, Fiévet, & Christophe, 2018; Kedar, Casasola, Lust, & Parmet, 2017; Kidd, White, & Aslin, 2011; Lew-Williams, 2017; Lew-Williams & Fernald, 2007; Lukyanenko & Fisher, 2016; Mani & Huettig, 2012; Reuter, Borovsky, & Lew-Williams, 2019; Waxman, 1999; Waxman, Lidz, Braun, & Lavin, 2009; Ylinen, Bosseler, Junttila, & Huotilainen, 2016; Yurovsky, Case, & Frank, 2017). While these studies do not provide evidence for prediction via simulation – specifically, they do not provide evidence that comprehenders take a speaker's visual perspective into account or simulate their word productions – they also do not contain evidence against prediction via simulation. They do suggest at least some capacity to use whatever linguistic or visual cues are available to efficiently interpret incoming speech. It is more likely, however, that children in these studies relied on phonological, morphosyntactic, and/or semantic associations between related words to generate real-time predictions, such as the use of informative verbs to anticipate upcoming nouns (Mani & Huettig, 2012).

Second language (L2) learners provide an interesting test case for understanding the emergence of prediction and the possibility of prediction via simulation more specifically. Much like children acquiring their first language, L2 adults could face difficulties in generating accurate predictions as they navigate referential contexts using their second language. While L2 adults have mature perspective-taking abilities, which are often needed for accurate prediction via simulation, they necessarily have less total experience with their L2, as compared to L1 adults, and may therefore have an overall reduced ability to generate predictions during language processing. Indeed, a number of prior findings indicate that L2 adults' predictions are attenuated and more variable than those of L1 adults, known as the Reduced Ability to Generate Expectations (RAGE) hypothesis (Grüter, Lew-Williams, & Fernald, 2012; Grüter, Rohde, & Schafer, 2017; Kaan, 2014; Lew-Williams, 2017; Lew-Williams & Fernald, 2010; Mitsugi & MacWhinney, 2016). For example, Lew-Williams and Fernald (2010) found that L2 Spanish-learning adults did not reliably use the grammatical gender of definite articles (i.e., *el* and *la*) to accurately anticipate upcoming referents in an eye-tracking task. However, much like the aforementioned developmental findings, existing evidence for prediction among L2 adults does not definitively determine whether they use association or simulation as a basis for generating those predictions. Although prior investigations for L1 children and L2 adults do not clearly differentiate which predictive mechanisms (i.e., association or simulation) may underlie the observed effects, they do suggest that everyday language experience is important for language learners of any age to

develop the abilities necessary for rapid and accurate predictions – whether via association, via simulation, or both.

The present study aimed to unite this previous literature on L1 adults, L1 children, and L2 adults in order to (1) further what is known about the mechanisms supporting prediction during real-time language processing, and (2) examine how differences in listeners' language experiences shape the diverse ways in which they could generate predictions. To evaluate prediction via simulation, we selected a particularly apt feature of language: spatial deixis, which includes words such as *this, that, these,* and *those.* Spatial deixis is a particularly useful test case for two reasons. First, *this* and *that* are singular, whereas *these* and *those* are plural. Deictic determiners convey both morphosyntactic and lexical semantic cues to number information, which in combination may support prediction, as found in prior research (Lew-Williams, 2017; Lew-Williams & Fernald, 2009; Lukyanenko & Fisher, 2016). Second, *this* and *these* typically indicate referents proximal to a speaker, whereas *that* and *those* typically indicate distal referents. Although the specific spatial interpretations of deictic determiners are not encoded morphosyntactically and can vary based on the particular referential context (Clark & Sengul, 1978; Diessel, 1999; Fillmore, 1997; see Levinson, 2004 for review), the lexical semantic information encoded in deictic determiners may in fact allow listeners to anticipate proximal or distal referents in real time. Critically, using spatial deixis to predict the referent's proximity to the speaker would require taking the speaker's perspective into account (i.e., simulation), because deictic words are anchored on the speaker's perspective. Whereas using deictic determiners to predict the *plurality* of the upcoming referent could be achieved via association, using deictic determiners to predict the *proximity* of the upcoming referent could only be achieved via simulation – taking the speaker's visual perspective into account and adjusting predictions accordingly.

In two experiments, we evaluated listeners' comprehension of spatial deixis with three groups of participants: native English-speaking adults, native English-learning 5-year-olds, and non-native English-learning adults. We refer to these groups as L1 adults, L1 children, and L2 adults, respectively, although there was notable heterogeneity within the L2 adult group. Five-year-old children were targeted for several reasons. Although the learning of deictic words may be slow and error-prone, prior evidence suggests that children reliably comprehend and produce deictic terms beginning around 4 years of age (Clark & Sengul, 1978; Tanz, 1980). Relatedly, previous findings indicate that children between the ages of 3 and 6 years old can use morphosyntactic number markings as a basis for predictions (Lew-Williams, 2017; Lew-Williams & Fernald, 2009; Lukyanenko & Fisher, 2016), therefore 5-year-olds could potentially use deictic determiners' number markings to anticipate the speaker's upcoming referent. However, evidence from referential communication tasks indicate that children's ability to accommodate a speaker's differing perspective is still developing at this age (Epley, Morewedge, & Keysar, 2004; Nadig, & Sedivy, 2002; Nilsen & Graham, 2009). Given that perspective-taking is one critical component of accurate prediction via simulation, it is possible that 5-year-olds may only be able to generate predictions via association, as claimed by Pickering and Garrod (2013). Based on prior developmental research, we expected that L1 children might use number marking cues during real-time language processing (which do not require perspective-taking), but fail to show reliable evidence of using proximity cues (which do require perspective-taking) as

a basis for generating predictions. Relatedly, we examined L2 adults based on Pickering and Garrod's (2013) claimed that comprehenders with less language experience – both L1 children and L2 adults – rely on the simpler association route for generating predictions. Prior findings do indicate more attenuated and variable predictions among L2 adults, as compared to L1 adults (Grüter et al., 2012; see Kaan, 2014 for a review), but these results do not address whether association or simulation might underlie the observed differences. The present study aims to examine nuanced, experience-based differences in comprehenders' predictive mechanisms (i.e., association vs. simulation). Specifically, by comparing predictive language processing among L1 adults, L1 children, and L2 adults, the present study aims to better understand how different predictive mechanisms may arise from listeners' varying language experiences.

Two eye-tracking tasks evaluated each group's comprehension of spatial deixis terms under conditions of referential complexity (Experiment 1) and reduced referential complexity (Experiment 2). Specifically, a cartoon speaker with the opposite perspective of the participants used various deictic determiners to refer to objects. Some referents were closer to the speaker, and some were closer to the participant; some were singular, and some were plural. The main hypothesis was that L1 adults would predict via simulation, taking the speaker's perspective into account during real-time language processing. In contrast, L1 children and L2 adults – theoretically due to their relative lack of experience comprehending and producing English sentences – may not predict via simulation. Importantly, L1 children and L2 adults may fail to do so for varying reasons: difficulty comprehending spatial deictic terms (Clark & Sengul, 1978; Tanz, 1980), difficulty taking the speaker's perspective and inhibiting their own, opposite perspective (Clark, 1992; Nilsen & Graham, 2009), or difficulty incorporating the surrounding linguistic context (Snedeker & Trueswell, 2004). More specifically, we predicted that L1 adults would rapidly use morphosyntactic number marking cues *and* lexical semantic proximity cues to predict the likely referent from the speaker's perspective, e.g., use *this* to look at a referent close to the speaker, or use *those* to look at referents close to themselves. Additionally, we expected that L1 children and L2 adults would successfully form predictions using morphosyntactic number marking cues, as in prior studies (Lew-Williams, 2017; Lew-Williams & Fernald, 2009; Lukyanenko & Fisher, 2016), but show relatively poorer abilities to predict using lexical semantic proximity cues. Together, by using spatial deixis as a lens to evaluate listeners' prediction abilities and by comparing three groups of participants with contrasting histories of language experience, these experiments further what is known about how prediction occurs during real-time processing.

Experiment 1

Method

Participants—Participants were 28 native English-speaking adults (11 male), 28 non-native English-learning adults (13 male), and 28 children (10 male) from monolingual English-speaking households. We refer to these groups as L1 adults, L2 adults, and L1 children, respectively. L1 adults and L2 adults were all members of the Princeton University campus community. L1 adults were 18 to 34 years old ($M = 20.54$ years, $SD = 4.23$ years),

L2 adults were 18 to 34 years old (M = 23.14 years, SD = 4.97 years), and L1 children were 60 to 71 months old (M = 64.9 months, SD = 3.8 months). L1 adults were significantly younger, on average, than L2 adults (t(52) = −2.11, p = 0.039). However, according to self-report measures, L2 adults had significantly fewer years of English exposure than L1 adults (L1 adults: M = 20.5, SD = 4.2, L2 adults: M = 16.7, SD = 4.9, t(53) = 3.14, p = 0.003). Note that 'L1' refers to any individual who learned English from birth, and 'L2' refers to any individual who learned English later. L2 adults varied in the age at which they reportedly began learning English (M = 6 years, SD = 3 years, range = 1–12 years). Two L2 adults in Experiment 1 learned English prior to age 3, and therefore may not meet traditional criteria for being an L2 learner, but including versus excluding them from analyses did not change any of the statistical analyses. Analyses that exclude these two participants are available in Supplementary Materials on the Open Science Framework (OSF), along with all deidentified data and additional descriptive information about L2 adults.

Given that we were interested in general experiential differences between L1 and L2 adults, and that we did not aim to evaluate a specific L1-L2 pairing, L2 adult participants represented a wide range of native languages: Bulgarian (2), Cantonese (6), German, Hebrew, Indonesian, Italian, Japanese, Kinyarwanda, Korean (4), Mandarin, Modern Greek, Nepalese, Portuguese, Russian (2), Spanish (2), Telugu, and Vietnamese. We tested but excluded one child participant (male, 61 months old) from all analyses due to the caregiver talking during the experiment. The research protocol was approved by the Princeton University Institutional Review Board (IRB record number 7117) and conformed to all guidelines for ethical treatment of participants.

Stimuli—Auditory stimuli were pre-recorded sentences, including: instructions, practice trials, test trials, and filler trials. Instructions introduced the computerized narrator of the task, the spatial context of the task, and the goal of the task: *"Hi, I'm Sally! I have a computer game for you. We'll see some things on this big, long table. Some things will be close to me, over here. Other things will be close to you, over there. I'll name something I see on the table. Try to find it with your eyes as fast as you can."* Two practice trials occurred immediately after instructions and further reinforced the spatial context of the task by providing a direct juxtaposition of two deictic demonstratives: *"Look at that happy cow over there. Now look at this happy cow over here. Look at this pretty horse over here. Now look at that pretty horse over there."* Test trials allowed us to evaluate whether participants could exploit deixis to predict an upcoming referent. Each test sentence was composed of a single command (*Look at*), one of five demonstratives (*the, this, that, these, those*), one of two adjectives (*beautiful, wonderful*), and a singular or plural target noun (*baby, doggy, kitty, turtle, apple, cookie, truck, bike*). Finally, filler trials included simple, affirmative statements (e.g., *"Wow! You're doing great!"*).

A female, native speaker of English recorded auditory stimuli, using child-directed intonation. We used Praat (Boersma & Weenink, 2017) to normalize the duration and intensity of the stimuli, such that each test sentence had a total duration of 2364 ms and a mean intensity of 65 dB. We aimed to assess whether listeners could use deictic determiners (e.g., *this*) to predict upcoming target nouns (e.g., *cookie*), so we also used Praat to identify the mean determiner onset (569 ms, range = 420 ms to 710 ms) and mean target noun onset

(1639 ms, range = 1429 ms to 1787 ms). Thus, on average, deixis onset occurred 1070 ms before the onset of the target noun.

Visual stimuli were a subset of images from a prior eye-tracking study (Lukyanenko & Fisher, 2016). Images included singular and plural versions of each target noun (e.g., one cookie, two cookies). The versions of the images were matched in size (for details on this approach, see Lukyanenko & Fisher, 2016). Visual stimuli also included an image of a female cartoon speaker. The speaker was positioned behind an image of a table that included depth perspective cues. Specifically, the table was wider at the bottom and narrower at the top of the image. The image of the speaker and table served as a backdrop for the four referent images (Figure 1).

During each test trial, four referents appeared. Two of the referents were plural, and two were singular. Two of the referents were proximal to the speaker, and two were distal to the speaker. Each trial included a plural referent and a singular referent proximal to the speaker, and a plural referent and a singular referent distal to the speaker, such that plurality and proximity were not conflated. Referents were visible for 2 seconds prior to the onset of the auditory stimuli.

Trials appeared in one of four quasi-randomized orders. Each order included instruction sentences, two practice trials, 32 test trials (16 with a deictic sentence and 16 with a neutral sentence), and four filler trials. Filler trials occurred every eight trials. Target side (left, right of the speaker), target plurality (singular, plural), and target proximity (proximal, distal to the speaker) were counterbalanced for each target noun. Target side, target plurality, and target proximity did not repeat for more than four consecutive trials. Visual stimuli, auditory stimuli, and experimental designs are available on the Open Science Framework (OSF).

Procedure—The experiment took place at the Princeton Baby Lab, in a sound-attenuated study room. The experimenter sat opposite the participant. Participants sat in a chair, approximately 60 cm from the eye-tracker. Child participants sat in a booster seat. In order for the eye-tracker to measure eye movements, participants wore a small target sticker on their face. The experimenter used EyeLink Experiment Builder software (SR Research, Mississauga, Ontario, Canada) and controlled the task from a Mac host computer. The experimenter first calibrated the eye-tracker for each participant using a standard 5-point calibration procedure. Participants listened to pre-recorded task instructions, then completed the eye-tracking task. Throughout the task, participants viewed stimuli on a 17-inch LCD monitor. An EyeLink 1000 Plus remote eye-tracker, sampling at a rate of 500 Hz, recorded participants' eye movements. The total duration of the eye-tracking task, including calibration, was approximately 4 minutes. Immediately following the eye-tracking task, L2 adults completed a questionnaire about their language background (see Appendix) in order to explore possible relations between prediction abilities and self-reported language proficiency.

Results

During the experiment, the eye tracker automatically recorded participants' fixations every 2 ms (500 Hz). We analyzed samples recorded within a 400×200 pixel area surrounding each

visual referent and eliminated any samples that were outside of these visual areas of interest (2,973,231 of 8,420,660 samples, 34%) prior to aggregating data within 100-ms time-bins. Further inspection of the data indicated high quality: We evaluated the number of time-bins that each participant contributed per trial, and found that, on average, participants provided data for 77% of time-bins (M = 12.5 bins, SD = 4.43 bins). If a participant did not contribute *any* data for a trial, then that trial was necessarily excluded from analyses. However, missing trials were rare. Participants contributed data for the vast majority of trials (M = 31 trials, SD = 1.17 trials, range = 26–32 trials), contributing data for 2638 of 2688 total trials (98%). That is, only 2% of trials were necessarily excluded due to missing data. We used R software (version 3.6.0) for all analyses. Deidentified data and R code for reproducible analyses are available on the Open Science Framework (OSF).

We completed three main analyses to evaluate listeners' task performance. First, we analyzed participants' looking behaviors during practice trials. If participants understood the spatial context of the task, then we expected them to reliably identify the target referent during practice trials. Next, we evaluated participants' looking behaviors during singular and plural deictic test sentences (comparing *this/that* to *these/those*). We then evaluated participants' looking behaviors during proximal and distal deictic test sentences (comparing *this/these* to *that/those*).[1] If participants use the plurality and proximity information conveyed by deictic determiners to predict the speaker's referent, then we expected to observe significant differences in looking behaviors before the onset of the target noun, indicating that participants launched anticipatory eye movements in response to the deictic determiners.

We used mixed-effects logistic regression models and cluster-based permutation analyses, detailed below, to evaluate participants' looking behaviors during deictic test sentences. Mixed-effects models are commonly used to analyze eye-tracking data (for reviews see: Barr, 2008; Barr, Levy, Scheepers, & Tily, 2013; Huettig, Rommers, & Meyer, 2011), and simultaneously account for fixed effects (i.e., variance attributable to the experimental conditions) and random effects (i.e., variance attributable to particular subjects and items). However, mixed-effects model results (namely, effects of time as a factor) must be interpreted cautiously because data from sequential time points are not independent. For example, where a participant is looking at 100 ms is largely dependent on where they were looking at 0 ms. Cluster-based permutation analyses can address this limitation and provide a useful analytical follow-up for mixed-effects models. These nonparametric analyses are commonly used for analyzing neurophysiological time course data (see Maris & Oostenveld, 2007 for a review), and are becoming increasingly common for analyzing eye-tracking data as well (Borovsky, 2017; Borovsky et al., 2015; Borovsky et al., 2016; Chan et al., 2018; Dautriche, Swingley, & Christophe, 2015; Hahn, Snedeker, & Rabagliati, 2015; Oakes et al., 2013; Reuter et al. 2021; Von Holzen & Mani, 2012; Wittenberg, Khan, & Snedeker, 2017).

---

[1] To evaluate potential timing differences between the deictic conditions being compared, we analyzed the time duration between the onset of the determiner (e.g., *this*) and the onset of the target noun (e.g., *cookie*) for each comparison with two-tailed, paired-sample *t*-tests. Comparing singular deictic sentences (*this/that*) and plural deictic sentences (*these/those*), we found no significant difference in durations between determiner onset and target noun onset (t(15) = 0.58, p = 0.571). Comparing proximal deictic sentences (*this/these*) and distal deictic sentences (*that/those*), we again found no significant timing differences across conditions (t(15) = 0.42, p = 0.682). Thus, results indicate that participants had approximately equal time to generate predictions across conditions.

Together, mixed-effects models and cluster-based permutation analyses aid in evaluating whether there may be differences in the time course of prediction among L1 adults, L1 children, and L2 adults.

Practice Trials—We first confirmed that participants understood the spatial context of the eye-tracking task by analyzing their looking behavior during practice trials. Practice trials used deixis contrastively and emphasized the proximity information encoded in deictic terms by pairing the proximal and distal deictic terms with "over here" and "over there" respectively. We analyzed participants' proportion of target looks during a time window from 200 ms after the exact onset of the deictic determiner to 2000 ms after the exact onset of the target noun using one-tailed one-sample t-tests to compare target looks to chance performance (50%). We found that all groups reliably looked to the target referent during practice trials (L1 adults: $t(27) = 11.35$, $p < 0.001$, Cohen's $d = 2.14$; L1 children: $t(27) = 3.75$, $p < 0.001$, Cohen's $d = 0.71$; L2 adults: $t(27) = 9.27$, $p < 0.001$, Cohen's $d = 1.75$). Findings therefore indicate participants reliably comprehended deictic terms when used contrastively. Importantly, these findings indicated that participants understood the spatial context of the eye-tracking task, which imitated a 3-dimensional conversational setting with 2-dimensional images, with some objects proximal to the speaker and some distal to the speaker.

Deictic Test Trials—In order to assess how participants used deictic determiners (e.g., *this*) to predict upcoming nouns (e.g., *cookie*), we analyzed participants' looks to referents during a time window from 1000 ms before target noun onset to 500 ms after noun onset. If participants use deictic determiners to predict the speaker's upcoming referent, we expected that effects would emerge at some point during this time. Importantly, if participants use deictic determiners to anticipate the plurality and proximity of the upcoming referent, then we should expect to observe the emergence of effects *before* the onset of the target noun (0 ms), with consideration in our analyses of the time it takes (approximately 200 ms) to initiate an eye movement (Matin, Shao, & Boff, 1993). Although it is possible that later effects could also reflect prediction (i.e., listeners could be more efficient in comprehending the target noun if it is preceded by an informative deictic determiner), we defined prediction as eye movements initiated prior to noun onset, as is common in prior research (e.g., Kidd, White, & Aslin, 2011; Mani & Huettig, 2012; Reuter et al., 2019). Our figures that summarize results by group and by experiment (Figure 2, Figure 3, and Figure 5) reveal how participants' looking behaviors changed over time, with 0 representing the exact onset of the target noun. Time measures are not offset by 200 ms to account for the time it takes to launch an eye movement.

Deictic Test Trials: Plurality—We first evaluated whether or not participants used deixis to predict the *plurality* of the referent. We analyzed listeners' proportion of looks to plural referents for singular deictic sentences (e.g., *this/that cookie)* and plural deictic sentences (e.g., *these/those cookies*) with a mixed-effects logistic regression model, using the lme4 package (version 1.1–21, Bates, Maechler, Bolker & Walker, 2015) and the lmerTest package (version 3.1–0, Kuznetsova, Brockhoff, & Christensen, 2017). The model included fixed effects for language group (treatment-coded contrasts: L1 adults, L1 children, L2

adults), condition (treatment-coded contrasts: plural terms = 0, singular terms = 1) and time (100-ms bins, −1000 to 500 ms from noun onset) and their interactions. The model also included random intercepts for subjects and items (Barr, Levy, Scheepers, & Tily, 2013).

As can be seen in Figure 2, model results (with L1 adults as the reference group) revealed a significant interaction of condition and time, indicating that L1 adults increasingly looked to the appropriate plural and singular referents over time ($\beta = -0.96$, $z = -21.04$, $p < 0.001$). Importantly, as illustrated by Figure 2, results revealed three-way interactions of language group, condition, and time, indicating that the interaction between condition and time for L1 children was more attenuated than that of L1 adults ($\beta = 0.48$, $z = 7.46$, $p < 0.001$) and that the interaction effect for L2 adults was likewise more attenuated than that of L1 adults' ($\beta = 0.29$, $z = 4.54$, $p < 0.001$). Together, results suggest that L1 adults, L1 children, and L2 adults differed in their patterns of looking behavior during plural and singular deictic sentences (Figure 2).

Figure 2 also conveys results from cluster-based permutation analyses (Maris, & Oostenveld, 2007; Wittenberg, Khan, & Snedeker, 2017). For these analyses, we calculated participants' mean proportion of looks to plural referents within each 100-ms time bin and performed a log-odds transformation on these proportions (Barr, 2008). Next, for each 100-ms time bin, we conducted a linear regression analysis on the log-odds of looking to plural referents. We identified clusters of time bins, defined as 2 or more adjacent time bins with t-values greater than 1.6 – a somewhat conservative value that has been used in prior eye-tracking research (Wittenberg, Khan, & Snedeker, 2017) – and summed t-values within each cluster. We then permuted the data to create the null distribution: We randomly shuffled condition labels 1000 times for each time bin, sampling across all time bins, and repeated the cluster-finding procedure and summation of t-values with these permuted data. Finally, we calculated the p-value for each cluster, defined as the proportion of permuted cluster t-values that were greater than the observed cluster t-value. Findings revealed significant clusters for L1 adults (−500 to 500 ms, cluster $t = 91.82$, $p < 0.001$), L1 children (−400 to 500 ms, cluster $t = 46.37$, $p < 0.001$), and L2 adults (−800 to 500 ms, cluster $t = 77.20$, $p < 0.001$). The observed differences in participants' looking behavior given plural versus singular deictic sentences – and, critically, the emergence of the effect prior to the onset of the number-marked noun – suggest that all groups used the morphosyntactic number marking of deictic determiners to anticipate the plurality of upcoming referents (Figure 2).

Deictic Test Trials: Proximity—We next evaluated whether participants used deixis to predict the *proximity* of the referent, using a mixed-effects logistic regression model and cluster-based permutation analyses, repeating the above plurality analyses. Figure 3 illustrates results for proximity analyses. The regression model included fixed effects for language group (treatment-coded contrasts: L1 adults, L1 children, L2 adults), condition (treatment-coded contrasts: proximal terms = 0, distal terms = 1) and time (100-ms bins, −1000 to 500 ms from noun onset) as well as their interactions, and included random intercepts for subjects and items. As can be seen in Figure 3, model results (with L1 adults as the reference group) again revealed a significant interaction of condition and time for L1 adults, indicating that they increasingly looked to the appropriate proximal and distal referents over time ($\beta = -0.82$, $z = -18.27$, $p < 0.001$). Importantly, as illustrated by

Figure 3, results again revealed three-way interactions of language group, condition, and time, indicating that the interaction between condition and time for L1 children was more attenuated than that of L1 adults ($\beta = 0.53$, $z = 8.42$, $p < 0.001$) and that the interaction effect for L2 adults was likewise more attenuated than that of L1 adults ($\beta = 0.14$, $z = 2.19$, $p = 0.029$). Moreover, findings from the cluster-based permutation analyses revealed significant clusters for L1 adults (−100 to 500 ms, cluster $t = 50.68$, $p < 0.001$), L1 children (300 to 500 ms, cluster $t = 14.45$, $p < 0.001$), and L2 adults (200 to 500 ms, cluster $t = 32.09$, $p < 0.001$).

The results summarized by Figure 3 collectively suggest that groups varied in using the proximity information of deictic determiners to predict the spatial location of the upcoming target referent. Whereas L1 adults' condition-based differences in looking behavior diverged *before* noun onset, indicating anticipatory eye movements to the appropriate referents, L1 children and L2 adults' looking behavior only diverged *after* noun onset, suggesting that they did not reliably use proximity information conveyed by the deictic determiner to anticipate the spatial location of the upcoming referent (Figure 3). The effect for L2 adults begins at approximately 200 ms following noun onset, and therefore likely reflects processing of the target noun itself, given that saccades are estimated to take 200 ms to initiate (Matin, Shao, & Boff, 1993). Together, this pattern of results suggests L1 adults quickly and accurately exploited proximity information as a basis for their predictions, whereas L1 children and L2 adults may not have generated predictions based on proximity at all, or may have done so slowly, inconsistently, or inaccurately (Figure 3).

Neutral Test Trials—We also used cluster-based permutation analyses to evaluate participants' looking behavior during neutral sentences in the same manner as for deictic sentences. The neutral determiner "the" does not provide information about the plurality or spatial location of the upcoming noun. We therefore expected to observe significant effects *after* the onset of the noun (0 ms), indicating that participants identified the appropriate referent once it was named. Comparing looks to plural referents for singular and plural neutral sentences, findings confirmed significant clusters for L1 adults (200 to 1000 ms, cluster $t = 200.22$, $p < 0.001$), L1 children (400 to 1000 ms, cluster $t = 64.50$, $p < 0.001$), and L2 adults (200 to 1000 ms, cluster $t = 136.39$, $p < 0.001$). Comparing looks to proximal referents for proximal and distal neutral sentences, findings again confirmed significant clusters for L1 adults (300 to 1000 ms, cluster $t = 159.86$, $p < 0.001$), L1 children (200 to 1000 ms, cluster $t = 58.27$, $p < 0.001$), and L2 adults (200 to 1000 ms, cluster $t = 138.40$, $p < 0.001$). These findings indicate that participants, upon hearing the neutral determiner "the", oriented to the correct referent after it was named.

Language Questionnaire—Immediately after the eye-tracking task, L2 adults completed a questionnaire which included various questions about their language experience. For example, L2 adults reported the age at which they began learning English ($M = 6.4$ years, $SD = 3.2$ years, range = 1 to 12 years), and total years of exposure to English ($M = 16.7$, $SD = 4.9$, range = 7 to 25 years). L2 adults also reported their self-assessed proficiency in a number of domains using a scale (1 through 9) with 1 indicating low proficiency and 9

indicating high proficiency (Table 1). According to these self-report measures, L2 adults had a high level of English proficiency.

We conducted exploratory analyses to correlate L2 adults' looking behaviors during eye tracking with the questionnaire measures. Specifically, we quantified each L2 adult's prediction measures as a difference score, subtracting their proportion of target looks during neutral trials from their proportion of target looks during deictic trials during a time window from 1000 ms before target noun onset to 200 ms after target noun onset. Participants with larger difference scores were therefore those who were better able to use deictic determiners to rapidly and accurately predict the speaker's upcoming referent and generate anticipatory eye movements to the corresponding image. Target looks generated 200 ms or later after the target noun would reflect processing of the target noun itself, as it takes approximately 200 ms to generate a saccade (Matin, Shao, & Boff, 1993). We found that L2 adults' prediction measures were not significantly correlated with the age at which they began learning English ($r(26) = 0.23$, $p = 0.231$), their total years of learning English ($r(26) = -0.22$, $p = 0.255$), their total years of English classes ($r(26) = 0.14$, $p = 0.469$), or their self-reported proficiency in understanding English ($r(26) = 0.003$, $p = 0.778$). Additional descriptive results from the language questionnaire are included in Supplementary Materials on the Open Science Framework (OSF).

Discussion

Experiment 1 results suggest that only adults listening to their first language may be capable of prediction via simulation – taking the speaker's perspective into account to rapidly and accurately simulate the speaker's upcoming production (Pickering & Garrod, 2013). Specifically, findings indicate that only L1 adults used deictic determiners to predict the proximity of the speaker's referent. That is, they were more likely to look towards a proximal referent when they heard *"Look at this/these…"* as compared to when they heard *"Look at that/those…"*, indicating consideration of the speaker's perspective (opposite from their own) to predict the spatial location of the referent. Importantly, the proximity effect emerged before L1 adults could have processed the target noun (e.g., cookie). That is, significant clusters emerged before 0 ms, indicating that L1 adults used the proximity information of deictic determiners to look towards proximal/distal referents before the target was identified. In contrast, significant effects for L1 children and L2 adults only emerged after the target was identified. L1 adults therefore showed robust evidence for prediction, while L1 children and L2 adults did not. Importantly, this conclusion is not based on the fact that significant clusters simply emerged earlier for L1 adults than for L1 children and L2 adults. Rather, it is based on the presence vs. absence of significant clusters before noun onset. L1 children's and L2 adults' more attenuated condition effects (i.e., delayed or shallow differences between proximal and distal conditions) might reflect predictions which are absent, inefficient, inconsistent, or inaccurate.

The findings of Experiment 1 are in line with the view that language experience shapes how comprehenders generate predictions during language processing (Pickering & Garrod, 2013). Importantly, all groups were capable of navigating some of the referential complexities of the task; L1 adults, L1 children, and L2 adults all performed above chance

during practice trials. Moreover, all groups used the morphosyntactic number markings of spatial deictic terms to predict the plurality of the speaker's upcoming referent, with significant condition effects emerging before the onset of the target noun (0 ms) for all three groups. This converges with prior results indicating that children can use morphosyntactic number marking cues (i.e., *is* vs. *are*) to anticipate the plurality of upcoming referents (Lukyanenko & Fisher, 2016). However, only L1 adults were able to use the lexical semantic proximity information of deictic determiners to anticipate the proximity of the referent.

Why did L1 children and L2 adults fail to show evidence for use of the spatial information conveyed by deictic determiners? Several linguistic and/or cognitive processes might explain the observed pattern of results. One possibility is that L1 children and L2 adults focused only on plurality cues (either inadvertently or deliberately) as a basis for generating predictions and failed to consider proximity cues. Presumably, proximity cues are more semantically ambiguous: Whereas *this* consistently refers to a singular referent, the spatial location of *this* changes based on the conversational context (i.e., the speaker's visual perspective). L1 children and L2 adults may have relied on a relatively less ambiguous cue to meaning – plurality – to generate predictions. In line with this view, L2 adults' results suggest they very rapidly distinguished plural and singular deictic terms, with a significant effect emerging 300 ms before that of L1 adults. Thus, L1 children and L2 adults may have identified morphosyntactic number markings as an unambiguous predictive cue and failed to consider more semantically ambiguous proximity cues as an additional way to narrow the scope of reference.

A related explanation for the observed pattern of results is that L1 children and L2 adults had difficulty rapidly *integrating* the two cues to meaning conveyed by deictic terms during real-time processing. That is, L1 children and L2 adults may be capable of using plurality cues *or* proximity cues to identify the speaker's upcoming referent, but may not be able to rapidly and accurately combine these two sources of information to generate predictions. Integrating these two cues to meaning may have taxed L1 children's and L2 adults' cognitive resources, such as working memory, such that they were unable to meet these task demands (Ito, Corley, & Pickering, 2018). The pattern of results among practice trials and test trials is consistent with this view: When deixis was used contrastively in practice trials, listeners did not need to contend with additional morphosyntactic plurality information, because referents in practice trials were all singular.

In Experiment 2, we aimed to reduce the task demands to determine whether a simplified referential context would facilitate L1 children's and L2 adults' use of deictic determiners to predict the proximity of the speaker's upcoming referent. To do so, we changed the task design so that visual stimuli in Experiment 2 included only two visual referents per test trial, rather than four. Importantly, the two referents were always matched in plurality. If L1 children and L2 adults failed to predict the proximity of referents due to difficulties integrating plurality and proximity information, then this experimental design should facilitate their ability to use deixis to anticipate the spatial location of the speaker's upcoming referent.

Experiment 2

Method

Participants—Participants were 28 native English-speaking adults (10 male), 28 non-native English-learning adults (13 male), and 28 children (13 male) from monolingual English-speaking households. We refer to these groups as L1 adults, L2 adults, and L1 children, respectively. L1 adults and L2 adults were all members of the Princeton University campus community. Children were 60 to 71 months old ($M = 63.4$ months, $SD = 2.9$ months), L1 adults were 18 to 21 years old ($M = 19.32$ years, $SD = 1.22$ years), and L2 adults were 18 to 35 years old ($M = 22.07$ years, $SD = 4.22$ years). L1 adults were significantly younger, on average, than L2 adults ($t(31) = -3.31$, $p = 0.002$). However, L2 adults reported significantly fewer years of English exposure than L1 adults (L1 adults: $M = 19.3$, $SD = 1.2$, L2 adults: $M = 14.5$, $SD = 3.7$, $t(32) = 6.55$, $p < 0.001$). As in Experiment 1, L2 adults began learning English at quite different ages from one other, and all L2 adults began learning English at age 3 or later ($M = 8$ years, $SD = 4$ years, range = 3–16 years). Deidentified data and additional descriptive information about L2 adults are available on the Open Science Framework (OSF).

L2 adults' native languages were: Cantonese, Hebrew, Japanese (2), Korean (5), Mandarin (3), Norwegian, Portuguese, Punjabi (2), Russian (4), Spanish (6), Urdu, and Vietnamese. We tested but excluded one L2 adult participant (female, 18 years old) from all analyses due to a computer error. The Princeton University Institutional Review Board approved this research protocol (Language Learning: Sounds, Words, and Grammar; IRB record number 7117) and research conformed to all guidelines for ethical treatment of participants.

Stimuli—All auditory stimuli were identical to Experiment 1, including deictic trials and neutral trials. Critically, visual stimuli in Experiment 2 included only two visual referents per test trial. The two referents in each test trial were always matched in plurality, such that they were both singular or both plural (Figure 4).

Procedure—The procedure was identical to Experiment 1: The experiment took place in a sound-attenuated study room at the Princeton Baby Lab. The experimenter calibrated the eye tracker for each participant using a standard 5-point procedure. Participants listened to pre-recorded task instructions, then completed the eye-tracking task. The total duration of the eye-tracking task, including calibration, was approximately 4 minutes. Immediately after the eye-tracking task, L2 adults completed a questionnaire about their language background.

Results

As in Experiment 1, the EyeLink eye tracker automatically recorded participants' fixations every 2 ms (500 Hz). We analyzed samples recorded within a 400×200 pixel area surrounding each visual referent and eliminated any samples that were outside of these visual areas of interest (2,321,591 of 7,003,083 samples, 33%) prior to aggregating data within 100-ms time-bins. Further inspection indicated data quality on par with Experiment 1: We evaluated the number of time bins that each participant contributed per trial, and found that, on average, participants provided data for 78% of time-bins ($M = 12.7$ bins, SD

= 4.29 bins). Missing trials were rare, as in Experiment 1. Participants contributed data for the vast majority of trials in Experiment 2 (M = 31 trials, SD = 1.07 trials, range = 27–32 trials), for a total of 2641 out of 2688 trials (98%). Thus, only 2% of trials were excluded due to missing data. Deidentified data and R code for reproducible analyses are available on the Open Science Framework (OSF).

Practice Trials—As in Experiment 1, we first analyzed participants' looking behavior during practice trials to confirm that they understood the spatial context of the task. We analyzed participants' looking behavior (aggregated from 200 ms after determiner onset to 2000 ms after target noun onset) with one-tailed one-sample t-tests, as in Experiment 1, and again found that all groups reliably looked to the target referent during practice trials (L1 adults t(27) = 12.25, p < 0.001, Cohen's d = 2.32; L1 children t(27) = 2.65, p = 0.007, Cohen's d = 0.50; L2 adults t(27) = 5.21, p < 0.001, Cohen's d = 0.98). Findings therefore indicate participants reliably comprehended deictic terms when used contrastively and understood the spatial context of the task.

Deictic Test Trials: Proximity—We again evaluated whether participants used deixis to predict the *proximity* of the referent (e.g., using *this* to look at a referent closer to the speaker) when the visual context was simplified (i.e., 2 visual referents instead of 4 referents) and when the plurality of the deictic term could not differentiate which referent the speaker intended because the visual referents were either both singular or both plural. As in Experiment 1, we analyzed looks to proximal referents for proximal deictic sentences (e.g., *this/these cookie/cookies)* and distal deictic sentences (e.g., *that/those cookie/cookies*) with a mixed-effects logistic regression model and cluster-based permutation analyses. The regression model included fixed effects for language group (treatment-coded contrasts: L1 adults, L1 children, L2 adults), condition (treatment-coded contrasts: proximal terms = 0, distal terms = 1) and time (100-ms bins, –1000 to 500 ms from noun onset) as well as their interactions, and included random intercepts for subjects and items.

As in Experiment 1, model results (with L1 adults as the reference group) revealed a significant interaction of condition and time, indicating that L1 adults increasingly looked to the appropriate proximal and distal referents, over time (β = –1.10, z = –25.41, p < 0.001). Figure 5 conveys these findings. Importantly, as illustrated by Figure 5, results revealed three-way interactions of language group, condition, and time, indicating that the interaction between condition and time for L1 children was more attenuated than that of L1 adults (β = 0.64, z = 10.62, p < 0.001) and that the interaction effect for L2 adults was likewise more attenuated than that of L1 adults (β = 0.16, z = 2.71, p = 0.007).

Figure 5 also displays findings from cluster-based permutation analyses, which revealed significant clusters for L1 adults (–600 to 500 ms, cluster t = 125.71, p < 0.001), L1 children (300 to 500 ms, cluster t = 19.97, p < 0.001), and L2 adults (–400 to –100 ms, cluster t = 7.13, p = 0.002; 0 to 500 ms, cluster t = 54.97, p < 0.001). Given that it takes approximately 200 ms to launch an eye movement (Matin, Shao, & Boff, 1993) the observed significant clusters which emerged prior to noun onset (0 ms) indicate that L1 adults and L2 adults made anticipatory eye movements to the target image before they had time to process the target noun itself. Thus, findings from Experiment 2 indicate that L1 adults and L2 adults,

but not L1 children, used spatial information conveyed by deictic determiners to anticipate the proximity of the upcoming referent. Experiment 2 results also converge with Experiment 1 findings by indicating that groups varied in their ability to predict the spatial location of the upcoming target referent. Whereas L1 adults demonstrate a robust condition difference before the onset of the target noun, the delayed condition difference among L1 children and the early but shallow condition difference among L2 adults again suggest that L1 children's and L2 adults' predictions were absent, inefficient, inconsistent, or inaccurate as compared to L1 adults' predictions (Figure 5).[2]

Neutral Test Trials—As in Experiment 1, we also used a cluster-based permutation analysis to assess participants' looking behavior during neutral sentences. Comparing looks to proximal referents for proximal and distal neutral sentences, results confirmed significant clusters for L1 adults (200 to 1000 ms, cluster $t = 291.91$, $p < 0.001$), L1 children (400 to 1000 ms, cluster $t = 83.51$, $p < 0.001$), and L2 adults (200 to 1000 ms, cluster $t = 185.55$, $p < 0.001$). These data show that participants looked to the appropriate referents after they were named in neutral trials.

Language Questionnaire—As in Experiment 1, L2 adults completed a self-report questionnaire about their language experiences, which included questions about the age at which they began learning English ($M = 7.5$ years, $SD = 3.6$ years, range = 3 to 16 years), their total years of exposure to English ($M = 14.5$, $SD = 3.7$, range = 8 to 26 years), and their self-assessed proficiency in a number of domains (Table 2). L2 adults in Experiment 2 reported a high level of English proficiency.

We completed the same exploratory analyses as in Experiment 1 to correlate L2 adults' looking behaviors during eye tracking with the questionnaire measures. We found that L2 adults' prediction measures were not significantly correlated with the age they began learning English ($t(26) = -0.11$, $p = 0.581$), their total years of learning English ($t(26) = -0.02$, $p = 0.935$), their total years of English classes ($t(26) = -0.14$, $p = 0.464$), or their self-reported proficiency in understanding English ($t(26) = 0.17$, $p = 0.401$). Additional descriptive results from the language questionnaire are included in Supplementary Materials on the Open Science Framework (OSF).

General Discussion

Despite a growing number of findings indicating that comprehenders generate predictions during language processing (for review see Kutas, DeLong, & Smith, 2011), the mechanisms underlying prediction and their developmental trajectories remain uncertain. One proposal by Pickering and Garrod (2013) includes two routes for prediction: *simulation* and *association*. According to this view, comprehenders can use production mechanisms (simulation) or comprehension mechanisms (association) to pre-activate upcoming representations. Importantly, the simulation route may only be accessible for

---

[2]·Participants had a general bias to look toward proximal referents in both experiments. This pattern of looking behavior is most likely due to the fact that the speaker, by definition, was located near the proximal referents (e.g., Figure 4). Given the limited number of referents to look at within the visual display, participants sometimes looked to the speaker, and this may have automatically biased them to look more toward proximal referents.

experienced, native comprehenders, whereas those with more limited language experience may rely on the association route (Pickering & Garrod, 2013; Pickering & Gambi, 2018). To evaluate these claims, the present study assessed how listeners with varied language experiences – L1 adults, L1 children, and L2 adults – generate predictions during language processing. Specifically, we used spatial deictic terms – *this, that, these,* and *those* – to determine whether listeners can generate predictions that take into account a speaker's perspective, which is a key feature of prediction via simulation. We hypothesized that all three groups of participants would use number information conveyed by deictic determiners to more efficiently process incoming sentences, but that only L1 adults would successfully take the speaker's spatial perspective into account and exploit the spatial information conveyed by deictic determiners. That is, we expected that L1 children and L2 adults – ostensibly due to their relative lack of English proficiency – would have difficulty doing so.

Findings supported these hypotheses. In Experiment 1, we found that only L1 adults used informative deictic determiners to anticipate the proximity of the speaker's referent, as evident from a significant effect before the target noun, but that L1 children and L2 adults only showed significant effects after the onset of the target noun. However, all groups used deixis to predict the plurality of the speaker's referent, as evident from significant effects before the target noun for all three groups. This pattern of results aligns with prior findings indicating that children can use morphosyntactic number marking cues (i.e., *el* and *los; is* and *are*) to predict an upcoming referent (Lew-Williams, 2017; Lew-Williams & Fernald, 2009; Lukyanenko & Fisher, 2016; Reuter et al. 2021). The significant plurality effect as well as participants' looking behaviors on practice trials also provide validation of the overall experimental design; that is, all three groups were able to navigate some aspects of its referential complexities. In Experiment 2, we used a simplified task with only two visual referents that were matched in plurality, which was intended to reduce task demands. Thus, only the proximity information of the determiner could provide disambiguating information for accurate prediction, such that listeners did not need to combine two sources of information (plurality and proximity) to generate predictions. We found that L1 adults and L2 adults used spatial information conveyed by deictic determiners to anticipate the proximity of the speaker's referent, as evident from significant effects prior to the target noun, but L1 children did not. In what follows, we consider potential causes for these observed group differences, discuss implications for theories of prediction, and pose questions for further investigation and theoretical refinement.

What might explain differences in performance among L1 adults, L1 children, L2 adults? There are a number of plausible explanations, which are not mutually exclusive. First, L1 children and L2 adults may have difficulty predicting the proximity of the speaker's referent due to difficulty with comprehending spatial deictic terms such as *this* vs. *that*. Prior findings indicate a slow and error-prone developmental trajectory for comprehending spatial deictic terms among L1 children, with reliable comprehension emerging at approximately four years of age (Clark & Sengul, 1978; see Tanz, 1980 for a review), and L2 adults may face similar challenges in processing deictic terms. However, this explanation is unlikely because L1 children and L2 adults correctly interpreted deictic terms during practice trials, accurately disambiguating *this cow* and *that cow* in the context of the spatial layout of the visual stimuli. That is, explicitly contrasting *this* and *that* with just one type of referent (as

opposed to multiple different referents on experimental deictic trials) allowed L1 children and L2 adults to exhibit successful comprehension of spatial deictic terms.

A second potential explanation for the observed group differences centers on perspective-taking. While children are generally aware that other people have distinct minds and perspectives before age 5 (Wellman, Cross, & Watson, 2010), L1 children may have had difficulty inhibiting their own visual perspective because the neurological systems supporting cognitive control are still developing at this age (Epley, Morewedge, & Keysar, 2004; Mazzarella, Ramsey, Conson, & Hamilton, 2013; Nadig, & Sedivy, 2002; Nilsen & Graham, 2009). Similarly, processing information in the second language may be more cognitively demanding for L2 comprehenders, which could result in a reduced ability to inhibit their own perspective and, in turn, to successfully use the simulation route for prediction. Experiment 2 findings provide some support for this possibility: L2 adults were able to take the speaker's perspective and predict via simulation when task demands were decreased. Moreover, prior findings indicate that L1 adults can have difficulty inhibiting their own perspective in tasks with increased cognitive load (Brown-Schmidt, 2009; Ito, Corley, & Pickering, 2017; Keysar, Barr, Balin, & Brauner, 2000; Keysar, Barr, & Horton, 1998; Keysar, Lin, & Bar, 2003; for review see Pickering & Gambi, 2018).

The observed group differences might also be explained by more general cognitive and/or linguistic factors. L1 children and L2 adults have difficulty integrating bottom-up and top-down information to rapidly and accurately simulate the speaker's upcoming production. That is, compared to L1 adults, they may have reduced speed and accuracy in processing bottom-up input and/or applying top-down, contextual constraints, including the speaker's perspective (Arnold, Brown-Schmidt, & Trueswell, 2007; Hurewitz et al., 2000; Kaan, 2014; Snedeker & Trueswell, 2004; Trueswell et al., 1999). Data for L2 adults in Experiment 2 provide some support for this possibility: When comprehension did not demand integration of both the plurality and proximity information of the deictic determiner, L2 adults succeeded in predicting the referent's proximity to the speaker. Further research is needed to tease apart these possibilities in order to form a more complete understanding of how comprehenders generate predictions during real-time processing.

A fourth, broad explanation for the observed pattern of results between L1 adults, L1 children, and L2 adults is that each group differed in the timing, quality, and quantity of their language learning experiences. There are obvious age-related differences between child and adult participants, but there are also experience-related differences that merit attention, particularly given that L2 adults had different patterns of predictive behaviors (relative to L1 adults) despite reporting high levels of English proficiency. A growing number of theoretical accounts and empirical findings suggest that comprehenders' learning histories shape the ways in which they process language and, perhaps necessarily, the ways in which they generate predictions during real-time processing (Chater, McCauley, & Christiansen, 2016; Dell & Chang, 2014; Grüter et al., 2012; Grüter et al., 2017; Kaan, 2014; Lew-Williams, 2017; Lew-Williams & Fernald, 2007; Martin et al., 2013; Mitsugi & MacWhinney, 2016; Pickering & Garrod, 2013).

What about L1 and L2 learners' histories with language may explain differences in the ability to exploit informative plurality and proximity cues in real-time processing? Although the present study did not systematically evaluate differences L1 and L2 learners' prior language experiences, we can point to some speculative answers to this question, based on prior research. A key feature that differentiates typical L1 and L2 environments is that L2 learners often learn in classroom settings, which differ from L1 learners' routine interactions with diverse interlocutors at home, at school, and in the community. Typical L1 learning in early childhood involves years of dynamic, moment-to-moment, multi-perspective exchange of information about events, objects, and people, with each person contributing their own perspective and generating their own action within a range of everyday contexts (Yurovsky, Smith, & Yu, 2012; Sheya & Smith, 2018). This experience is likely to yield rich exposure to deictic words as children and their caregivers engage with each other during play, book-reading, meals, and other contexts and routines. Indeed, some evidence suggests that deictic words are more common in child-directed speech than in adult-directed speech: In one study, 16% of child-directed utterances contained a deictic word, compared to only 2% of adult-directed utterances (Newport, Gleitman, & Gleitman, 1977). Similarly, in an exploratory evaluation of CHILDES data (MacWhinney, 2000), we found that 5-year-old children ($N = 39$) produced deictic terms (*this, that, these, those*) in 13% of their utterances, and adults conversing with 5-year-old children ($N = 109$) produced deictic terms in 17% of their utterances.[3] However, our null findings with L1 children suggest that this immersive experience does not automatically yield early efficiency in using spatial deictic words to predict upcoming referents, but instead requires substantial experience over many years, perhaps – as noted above – in combination with maturing abilities to engage in cognitive control. Relatedly, exposure to written language may support deixis-based prediction. L1 adults' and L1 children's experience with written language has been shown to influence the extent to which they generate predictions while processing spoken language (Huettig, Singh, & Mishra, 2011; Mishra, Singh, Pandey, & Huettig, 2012; Mani & Huettig, 2014).

In contrast to L1 learning environments, typical L2 learning environments are classroom-based and less immersive (Cazden, 1988; Lew-Williams, 2017; see National Academies of Sciences, Engineering, and Medicine, 2018). Conversations often involve specific prompts by a teacher or textbook, and learning is more likely to include memorization of words and grammatical constructions. These experiences would certainly include the use of spatial deictic words. However, even if frequency of exposure was comparable between L1 children and L2 adults, the nature of these exposures would be mostly distinct. L2 adults would be unlikely to hear deictic words embedded in the same range of natural contexts, and would be more likely to learn about these words in a top-down, grammar-based framework, which may drive differences in real-time language processing. Plurality is such a pervasive feature of language that all participant groups were able to exploit morphosyntactic number marking cues in real time, but the added perspective-taking demands involving spatial

---

[3.]This tentatively suggests that adults and children commonly use spatial deictic terms during everyday conversations. However, transcript data do not reveal whether deictic determiners were being used *spatially* (i.e., contrasting proximal and distal referents) as in the present study, nor do they indicate whether children and/or adults produced deictic determiners in contexts with shared, opposite, or orthogonal perspectives. Future investigations could use naturalistic video corpora to evaluate how caregivers and children use deictic determiners during everyday conversations. A project on natural use of spatial deixis in real-world referential contexts would be an excellent and ambitious complement to the present investigation.

deixis prevented L1 children and L2 adults from exploiting the same prediction routes as observed for L1 adults. Indeed, the ability to predict via simulation is thought to arise from spending several years comprehending language before you start producing longer sentences yourself; immediate practice with speaking, which is common in L2 learning, could have consequences for how learners integrate comprehension and production processes (Pickering & Garrod, 2013; Pickering & Gambi, 2018).

Thus, the present findings support a speculative view that quantitative and qualitative differences in how L1 and L2 learners gain experience listening to deictic words could incrementally shape their prediction mechanisms over time. This explanation aligns with Pickering and Garrod's (2013) proposal that comprehenders with limited language experience rely on comprehension-based mechanisms (i.e., association) whereas those with greater proficiency can additionally use production-based mechanisms (i.e., simulation). Prolonged, immersive language experiences may be necessary to attain higher levels of proficiency in processing and predicting language from multiple interlocutors' perspectives. Further research is needed to determine the details of day-to-day language use that support comprehenders' ability to generate efficient and accurate predictions about upcoming information, both for deictic words and other word forms.

By revealing experience-based differences in how listeners generate predictions, the present findings both support and refine existing theoretical accounts of language processing and acquisition. Findings provide empirical evidence which further validates proposed production-based versus comprehension-based mechanisms underlying prediction (Pickering & Garrod, 2013). That is, beyond indicating differences in *whether* listeners predict, the present findings reveal potential differences in *how* listeners predict. L1 adults can readily adjust their predictions based on a speaker's perspective (i.e., simulation), but in contrast, L1 children and L2 adults appear to have difficulty taking a speaker's perspective into account during real-time processing and instead primarily generate predictions which do not require contextual adjustments (i.e., association). Results also align with current theoretical views of L2 acquisition, in that L2 adults' prediction abilities may be more attenuated and variable as compared to L1 adults' (Pickering & Garrod, 2013; Grüter et al., 2012; Grüter et al., 2017). The present findings also broaden our understanding of how predictive mechanisms could shape the course of language acquisition (Dell & Chang, 2014) in that they point to association (rather than simulation) as a potential developmental mechanism. While prior correlational evidence suggests that children employ production-based, simulation mechanisms for prediction (Mani & Huettig, 2012), the present findings suggest that listeners with more limited language experience may instead rely on comprehension-based, association mechanisms.

Our findings raise a number of questions for further experimentation and theoretical refinement. An overarching question concerns the mechanisms underlying prediction, their developmental trajectories, and the role of language experience and cognitive factors in shaping them. If comprehenders with more limited experience rely on comprehension-based, association mechanisms, then when and how do they transition to additionally using production-based, simulation mechanisms to generate predictions (McCauley & Christiansen, 2013)? Why should learners use two semi-distinct routes for prediction that

may ultimately be redundant in terms of outcomes (Yoon & Brown-Schmidt, 2013)? A related question arises from Experiment 1 findings. Although only L1 adults predicted the proximity of the upcoming referent, all three groups of listeners were able to predict its plurality. How do comprehenders determine what sources of information to use as a basis for prediction during real-time language processing, and what circumstances allow for simulation beyond association (Trude, 2013)? By extension, what language processing experiences provide comprehenders with opportunities for learning via the back-propagation of prediction errors (Dell & Chang, 2014; Chang, Dell, & Bock, 2006; Elman, 1990)? It is unclear which predictive mechanisms (association, simulation, or both) support language acquisition, and further research is needed to systematically evaluate the developmental emergence of predictive language processing from infancy through adulthood. Likewise, learners' language abilities develop in tandem with their cognitive abilities, and further studies are needed to understand how learners' predictive mechanisms might be shaped by developmental changes in working memory and cognitive control (Mani & Huettig, 2013; Slevc & Novick, 2013). Although the present findings add to a growing body of research evaluating prediction, many questions remain. Addressing these questions, among others (see Huettig, 2015; Huettig & Mani, 2016; Rabagliati, Gambi, & Pickering, 2016), will aid in clarifying what role prediction may play in both language processing and acquisition.

Various limitations will need to be addressed to form a more complete understanding of how L1 and L2 comprehenders generate predictions during language processing. First, *not* observing prediction using behavioral measures is not definitive evidence for a lack of prediction. As compared to L1 adults' robust and accurate predictions in the present studies, L1 children's and L2 adults' predictions may have been absent, inefficient, inconsistent, or inaccurate. These possibilities are not mutually exclusive and may be difficult to definitively tease apart, but a combination of experimental approaches could be revealing. For example, future studies could combine behavioral and neuroimaging methods with high temporal resolution (e.g., eye-tracking and EEG) to attain multiple measures and provide a more complete evaluation of comprehenders' predictions during real-time processing. A related interpretational limitation concerns the small sample sizes per group, which leads to relatively low statistical power. A lack of statistical power is common in developmental investigations (Oakes, 2017) as well as in L2 investigations (Brysbaert, 2020). Although sample sizes for the present study were on par with similar research paradigms (Ito et al., 2018; Van Bergen & Flecken, 2017), replication and extension of the present findings will be important to further verify and build upon these results.

Another notable limitation stems from the intentional heterogeneity of the L2 adults in the present study. Namely, L2 adults had varied native languages (17 different languages in Experiment 1 and 12 different languages in Experiment 2). Some of L2 adults' native languages have different deictic demonstrative systems as compared to English. For example, Spanish terms denote three levels of proximity (e.g., *este*, *ese*, and *aquel*), whereas English terms denote two levels of proximity (e.g., *this* and *that*). Prior studies have documented cross-linguistic transfer effects among bilingual adults (Dussias, Valdés Kroff, Guzzardo Tamargo, & Gerfen, 2013; Jackson & Dussias, 2009; Lemhöfer, Spalek, & Schriefers, 2008; Sabourin, Stowe, & De Haan, 2006; but see Costa, Kovacic, Franck, & Caramazza, 2003), and it is possible that L2 adults' native languages shaped their

performance in the present study as well. In these first studies addressing listeners' real-time processing of deixis and prediction via simulation, we did not aim to compare different languages with varying patterns of morphosyntactic plurality cues and lexical semantic proximity cues. Doing so would certainly provide greater theoretical depth but would also require a range of robust, language-specific sample sizes; this is an important direction for future research on deixis. Similarly, L2 adults varied in their self-reported measures of English proficiency. Although L2 adults' proficiency measures were not significantly correlated with their prediction measures in the present study, future studies could more systematically evaluate potential links between the two by using a broader range of thorough, standardized proficiency measures. The present study may provide a valuable paradigm for evaluating cross-linguistic differences as well as individual differences.

The present study was also limited by an aspect of its design: Namely, the listener's perspective was always opposite from the speaker's perspective. This design was important for evaluating prediction via simulation (Pickering & Garrod, 2013): If the speaker and listener shared the same perspective, findings would not indicate whether the listener was taking the speaker's perspective or relying on their own, internal perspective as a basis for prediction. However, this design prevents us from teasing apart the specific factors that may contribute, individually or jointly, to the observed pattern of results. For example, future investigations could evaluate how L1 children and L2 adults interpret deictic determiners when they share the speakers' perspective or when their perspective is only somewhat rotated from the speakers'. Doing so would help to assess whether incorporating the speaker's opposite perspective during real-time processing influences the extent to which L1 children and L2 adults generate predictions using deixis. That is, it's possible that L1 children and L2 adults could better contend with simultaneous plurality cues and proximity cues (as in Experiment 1) if they did not have to take into account the speaker's opposite perspective.

Finally, findings from constrained experimental contexts must be interpreted with caution for a number of reasons. Participants (especially adults) may attempt to use explicit strategies in lab settings that could artificially inflate the observed effects and would not be tenable in real-world contexts. The present study instructed participants from all three groups to identify referents quickly, which could have provoked explicit strategies that are unlikely to occur in naturalistic conversations; nonetheless, L1 and L2 adults, who should have comparable abilities to engage in processing strategies, showed somewhat different performance in the experiment, suggesting that the instructions were not solely responsible for observed effects. In the future, investigations could provide no (or more minimal) task instructions to explore this possibility further. However, eliminating task instructions in lab-based investigations would not adequately address the broader issue of external validity. The external validity of results from constrained lab tasks is generally unknown (Huettig & Mani, 2016), and the simplified conversational context of the present study is unlikely to reflect the richness of comprehenders' typical day-to-day language use. The extent to which prediction occurs within the naturalistic speech and visual environment remains largely unexplored, although a growing number of studies have used more naturalistic experimental stimuli to evaluate adults' and children's prediction abilities (Andersson et al., 2011; Bögels, 2020; Coco et al., 2016; Reuter et al. 2021; Staub et al., 2012). Future research could

use head-mounted eye-tracking methods to evaluate how comprehenders navigate more naturalistic, 3-dimensional contexts, converse with interlocutors who have somewhat shared or somewhat different perspectives, and generate predictions based on combined social, spatial, and linguistic information.

In sum, the present study used two experiments with three groups of listeners (L1 adults, L1 children, and L2 adults) to explore experience-based variation in predictive language processing. The findings broaden our knowledge of the mechanisms underlying prediction, provide empirical support for current psycholinguistic theories (e.g., Pickering & Garrod, 2013), and raise promising avenues for continued theoretical refinement. Further research is needed to understand the interconnected set of factors underlying the observed group differences and, more broadly, to understand how comprehenders' day-to-day language use with other people continuously shapes the ways in which they process and predict information.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Appendix

Language Questionnaire

Do you consider yourself to be a native speaker of English?

If no, what do you consider to be your native language(s)?

What language(s) was/were spoken in your household before you started college?

At what age did you start learning English? (If you started learning from birth, please write 0.)

What percent of your interactions with your family and friends was in English...

before elementary school?

during elementary school?

during middle school?

during high school?

during college / after high school?

during this past year?

during this past week?

If you took English Language Learner (ELL) classes, please indicate how many years of classes you took...

during elementary school

during middle school

during high school

during college / after high school

Do you watch television/movies and/or listen to music in English? (Never, Sometimes, Often)

How would you rate your proficiency in speaking English? (1 to 9)

How would you rate your proficiency in understanding English? (1 to 9)

How would you rate your proficiency in reading English? (1 to 9)

How would you rate your proficiency in writing English? (1 to 9)

How would you rate your accent when speaking English? (1 to 9)

How would you rate your comfort when speaking English? (1 to 9)

References

Andersson R, Ferreira F, & Henderson JM (2011). I see what you're saying: The integration of complex speech and scenes during language comprehension. Acta Psychologica, 137, 208–216. 10.1016/j.actpsy.2011.01.007 [PubMed: 21303711]

Arnold JE, Brown-Schmidt S, & Trueswell J. (2007). Children's use of gender and order-of-mention during pronoun comprehension. Language and Cognitive Processes, 22(4), 527–565. 10.1080/01690960600845950

Barr DJ (2008). Analyzing "visual world" eyetracking data using multilevel logistic regression. Journal of Memory and Language, 59(4), 457–474. 10.1016/j.jml.2007.09.002

Barr DJ, Levy R, Scheepers C, & Tily HJ (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. Journal of Memory and Language, 68, 255–278. 10.1016/j.jml.2012.11.001

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

Boersma P, & Weenink D. (2017). Praat: doing phonetics by computer [Computer program]. Version 6.0.19, retrieved 13 June 2016 from http://www.praat.org/

Bögels S. (2020). Neural correlates of turn-taking in the wild: Response planning starts early in free interviews. Cognition, 203, 104347. 10.1016/j.cognition.2020.104347 [PubMed: 32563734]

Borovsky A. (2017). The amount and structure of prior event experience affects anticipatory sentence interpretation. Language, Cognition and Neuroscience, 32(2), 190–204. 10.1080/23273798.2016.1238494

Borovsky A, & Creel S. (2014). Children and adults integrate talker and verb information in online processing. Developmental Psychology, 50(5), 1600–1613. 10.1037/a0035591 [PubMed: 24611671]

Borovsky A, Ellis EM, Evans JL, & Elman JL (2015). Lexical leverage: Category knowledge boosts real-time novel word recognition in 2-year-olds. Developmental Science, 6, 918–932. 10.1111/desc.12343

Borovsky A, Ellis EM, Evans JL, & Elman JL (2016). Semantic Structure in Vocabulary Knowledge Interacts With Lexical and Sentence Processing in Infancy. Child Development, 87(6), 1893–1908. 10.1111/cdev.12554 [PubMed: 27302575]

Borovsky A, Elman JL, & Fernald A. (2012). Knowing a lot for one's age: Vocabulary skill and not age is associated with anticipatory incremental sentence interpretation in children and adults. Journal of Experimental Child Psychology, 112(4), 417–436. https://dx.doi.org/10.1016%2Fj.jecp.2012.01.005 [PubMed: 22632758]

Brysbaert M. (2020). Power considerations in bilingualism research: Time to step up our game. Bilingualism: Language and Cognition, 1–18. 10.31234/osf.io/92npz

Cazden CB (1988). Classroom discourse: The language of teaching and learning. Portsmouth, NH: Heinemann.

Chan A, Yang W, Chang F, & Kidd E. (2018). Four-year-old Cantonese-speaking children's on-line processing of relative clauses: A permutation analysis. Journal of Child Language, 45(1), 174–203. 10.1017/S0305000917000198 [PubMed: 28606194]

Chang F, Dell GS, & Bock K. (2006). Becoming syntactic. Psychological Review, 113(2), 234–72. 10.1037/0033-295X.113.2.234 [PubMed: 16637761]

Christiansen MH, & Chater N. (2016). The Now-or-Never Bottleneck: A fundamental constraint on language. The Behavioral and Brain Sciences, 39, E62. 10.1017/S0140525X1500031X [PubMed: 25869618]

Clark EV, & Sengul CJ (1978). Strategies in the acquisition of deixis. Journal of Child Language, 5(3), 457–475. 10.1017/S0305000900002099

Clark HH (1992). Arenas of Language Use. Chicago: University of Chicago Press.

Coco MI, Keller F, & Malcolm GL (2016). Anticipation in real-world scenes: The role of visual context and visual memory. Cognitive Science, 40(8), 1995–2024. 10.1111/cogs.12313 [PubMed: 26519097]

Costa A, Kovacic D, Franck J, & Caramazza A. (2003). On the autonomy of the grammatical gender systems of the two languages of a bilingual. Bilingualism: Language and Cognition, 6(3), 181–200. 10.1017/s1366728903001123

Dautriche I, Swingley D, & Christophe A. (2015). Learning novel phonological neighbors: Syntactic category matters. Cognition, 143, 77–86. 10.1016/j.cognition.2015.06.003 [PubMed: 26114905]

Dussias PE, Valdés Kroff JR, Guzzardo Tamargo RE, & Gerfen C. (2013). When gender and looking go hand in hand. Studies in Second Language Acquisition, 35(2), 353–387. 10.1017/S0272263112000915

Epley N, Morewedge CK, & Keysar B. (2004). Perspective taking in children and adults: Equivalent egocentrism but differential correction. Journal of Experimental Social Psychology, 40(6), 760–768. 10.1016/j.jesp.2004.02.002

Fernald A, Zangl R, Portillo AL, & Marchman VA (2008). Looking while listening: Using eye movements to monitor spoken language comprehension by infants and young children. In Sekerina IA, Fernandez EM, & Clahsen H.(Eds.), Developmental Psycholinguistics: Online Methods in Children's Language Processing (pp. 97–135). Amsterdam, Netherlands: John Benjamins Publishing Company.

Fillmore CJ (1997). Lectures on Deixis. Stanford, CA: CSLI (Original work published 1977).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Grüter T, Lew-Williams C, & Fernald A. (2012). Grammatical gender in L2: A production or a real-time processing problem? Second Language Research, 28(2), 191–215. 10.1177/0267658312437990 [PubMed: 30319164]

Grüter T, Rohde H, & Schafer AJ (2017). Coreference and discourse coherence in L2: The roles of grammatical aspect and referential form. Linguistic Approaches to Bilingualism, 7, 199–229. 10.1075/lab.15011.gru

Hahn N, Snedeker J, and Rabagliati H. (2015). Rapid linguistic ambiguity resolution in young children with autism spectrum disorder: eye tracking evidence for the limits of weak central coherence. Autism Research, 8, 717–726. [PubMed: 25820816]

Havron N, de Carvalho A, Fiévet AC, & Christophe A. (2018). Three- to four-year-old children rapidly adapt their predictions and use them to learn novel word meanings. Child Development, 90, 1–9. 10.1111/cdev.13113

Huettig F, & Mani N. (2016). Is prediction necessary to understand language? Probably not. Language, Cognition and Neuroscience, 31(1), 19–31. 10.1080/23273798.2015.1072223

Huettig F, Rommers J, & Meyer AS (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. Acta Psychologica, 137, 151–171. 10.1016/j.actpsy.2010.11.003 [PubMed: 21288498]

Huettig F, Singh N, & Mishra RK (2011). Language-mediated visual orienting behavior in low and high literates. Frontiers in Psychology, 2, 1–14. 10.3389/fpsyg.2011.00285 [PubMed: 21713130]

Hurewitz F, Brown-Schmidt S, Thorpe K, Gleitman L, & Trueswell J. (2000). One frog, two frog, red frog, blue frog: Factors affecting children's syntactic choices in production and comprehension. Journal of Psycholinguistic Research, 29, 597–626. 10.1023/A:1026468209238 [PubMed: 11196065]

Ito A, Corley M, & Pickering MJ (2017). A cognitive load delays predictive eye movements similarly during L1 and L2 comprehension. Bilingualism, 1, 1–14. 10.1017/S1366728917000050

Ito A, Martin AE, & Nieuwland MS (2017). On predicting form and meaning in a second language. Journal of Experimental Psychology: Learning Memory and Cognition, 43(4), 635–652. 10.1037/xlm0000315

Ito A, Pickering MJ, & Corley M. (2018). Investigating the time-course of phonological prediction in native and non-native speakers of English: A visual world eye-tracking study. Journal of Memory and Language, 98, 1–11. 10.1016/j.jml.2017.09.002

Jackson CN, & Dussias PE (2009). Cross-linguistic differences and their impact on L2 sentence processing. Bilingualism, 12(1), 65–82. 10.1017/S1366728908003908

Kaan E. (2014). Predictive sentence processing in L2 and L1: What is different? Linguistic Approaches to Bilingualism, 4, 257–282. 10.1075/lab.4.2.05kaa

Kedar Y, Casasola M, Lust B, & Parmet Y. (2017). Little words, big impact: Determiners begin to bootstrap reference by 12 months. Language Learning and Development, 13(3), 317–334. 10.1080/15475441.2017.1283229

Keysar B, Barr DJ, & Horton WS (1998). The egocentric basis of language use: Insights from a processing approach. Current Directions in Psychological Science, 7(2), 46–49. 10.1111/1467-8721.ep13175613

Keysar B, Barr DJ, Balin JA, & Brauner JS (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension. Psychological Science, 11(1), 32–38. 10.1111/1467-9280.00211 [PubMed: 11228840]

Keysar B, Lin S, & Barr DJ (2003). Limits on theory of mind use in adults. Cognition, 89(1), 25–41. 10.1016/s0010-0277(03)00064-7 [PubMed: 12893123]

Kidd C, White KS, & Aslin RN (2011). Toddlers use speech disfluencies to predict speakers' referential intentions. Developmental Science, 14(4), 925–934. https://dx.doi.org/10.1111%2Fj.1467-7687.2011.01049.x [PubMed: 21676111]

Kutas M, DeLong KA, & Smith NJ (2011). A look around at what lies ahead: Prediction and predictability in language processing. In Bar M. (Ed.) Predictions in the Brain: Using Our Past to Generate a Future (pp. 190–207). Oxford: Oxford University Press.

Kuznetsova A, Brockhoff PB, & Christensen RHB (2017). lmerTest Package: Tests in Linear Mixed Effects Models. Journal of Statistical Software, 82(13), 1–26. 10.18637/jss.v082.i13

Lemhöfer K, Spalek K, & Schriefers H. (2008). Cross-language effects of grammatical gender in bilingual word recognition and production. Journal of Memory and Language, 59(3), 312–330. 10.1016/j.jml.2008.06.005

Levinson S. (2004). Deixis. In Horn L, & Ward G. (Eds.), The Handbook of Pragmatics (pp. 97–121). Oxford: Blackwell.

Lew-Williams C. (2017). Specific referential contexts shape efficiency in second language processing: Three eye-tracking experiments with 6- and 10-year-old children in Spanish immersion schools. Annual Review of Applied Linguistics, 37, 128–147. 10.1017/s0267190517000101

Lew-Williams C, & Fernald A. (2007). Young children learning Spanish make rapid use of grammatical gender in spoken word recognition. Psychological Science, 18(3), 193–198. https://dx.doi.org/10.1111%2Fj.1467-9280.2007.01871.x [PubMed: 17444909]

Lew-Williams C, & Fernald A. (2009). Fluency in using morphosyntactic cues to establish reference: How do native and non-native speakers differ? Proceedings of the 33rd Annual Boston University Conference on Language Development.

Lew-Williams C, & Fernald A. (2010). Real-time processing of gender-marked articles by native and non-native Spanish speakers. Journal of Memory and Language, 63(4), 447–464. 10.1016/j.jml.2010.07.003 [PubMed: 21076648]

Lukyanenko C, & Fisher C. (2016). Where are the cookies? Two- and three-year-olds use number-marked verbs to anticipate upcoming nouns. Cognition, 146, 349–370. 10.1016/j.cognition.2015.10.012 [PubMed: 26513355]

MacWhinney B. (2000). The CHILDES Project: Tools for analyzing talk. Mahwah, NJ: Lawrence Erlbaum Associates.

Mani N, & Huettig F. (2012). Prediction during language processing is a piece of cake—But only for skilled producers. Journal of Experimental Psychology: Human Perception and Performance, 38, 843–847. 10.1037/a0029284 [PubMed: 22774799]

Mani N, & Huettig F. (2013). Towards a complete multiple-mechanism account of predictive language processing. Behavioral and Brain Sciences, 36, 365–366.

Maris E, & Oostenveld R. (2007). Nonparametric statistical testing of EEG- and MEG-data. Journal of Neuroscience Methods, 164(1), 177–190. 10.1016/j.jneumeth.2007.03.024 [PubMed: 17517438]

Martin CD, Thierry G, Kuipers JR, Boutonnet B, Foucart A, & Costa A. (2013). Bilinguals reading in their second language do not predict upcoming words as native readers do. Journal of Memory and Language, 69(4), 574–588. 10.1016/j.jml.2013.08.001

Matin E, Shao K. & Boff K. (1993). Saccadic overhead: Information processing time with and without saccades. Perception & Psychophysics, 53, 372–380. [PubMed: 8483701]

Mazzarella E, Ramsey R, Conson M, & Hamilton A. (2013). Brain systems for visual perspective taking and action perception. Social Neuroscience, 8(3), 248–267. 10.1080/17470919.2012.761160 [PubMed: 23350907]

McCauley S, & Christiansen M. (2013). Toward a unified account of comprehension and production in language development. Behavioral and Brain Sciences, 36(4), 366–367. 10.1017/S0140525X12002658

Mishra RK, Singh N, Pandey A, and Huettig F.. (2012). Spoken language-mediated anticipatory eye movements are modulated by reading ability: Evidence from Indian low and high literates. Journal of Eye Movement Research 5(1). 1–10. 10.16910/jemr.5.1.3

Mitsugi S, & MacWhinney B. (2016). The use of case marking for predictive processing in second language Japanese. Bilingualism: Language and Cognition, 19(1), 19–35. 10.1017/S1366728914000881

Nadig AS, & Sedivy JC (2002). Evidence of perspective-taking constraints in children's on-line reference resolution. Psychological Science 13(4), 329–36. 10.1111/j.0956-7976.2002.00460.x [PubMed: 12137135]

National Academies of Sciences, Engineering, and Medicine (2018). How people learn II: Learners, Contexts, and Cultures. Washington, DC: The National Academies Press. 10.17226/24783

Newport EL, Gleitman H, & Gleitman LR (1977). Mother I'd rather do it myself: Some effects and non-effects of maternal speech style. In Snow C. & Ferguson C. (Eds.), Talking to children: Language input and acquisition (pp. 109–149). New York: Cambridge University Press.

Nilsen ES, & Graham SA (2009). The relations between children's communicative perspective-taking and executive functioning. Cognitive Psychology, 58(2), 220–249. 10.1016/j.cogpsych.2008.07.002 [PubMed: 18809176]

Oakes LM (2017). Sample size, statistical power, and false conclusions in infant looking-time research. Infancy, 22(4), 436–469. 10.1111/infa.12186 [PubMed: 28966558]

Oakes LM, Baumgartner HA, Barrett FS, Messenger IM, & Luck SJ (2013). Developmental changes in visual short-term memory in infancy: Evidence from eye-tracking. Frontiers in Psychology, 4(OCT), 1–13. 10.3389/fpsyg.2013.00697 [PubMed: 23382719]

Pickering MJ, & Gambi C. (2018). Predicting while comprehending language: A theory and review. Psychological Bulletin, 144(10), 1002–1044. 10.1037/bul0000158 [PubMed: 29952584]

Pickering MJ, & Garrod S. (2007). Do people use language production to make predictions during comprehension? Trends in Cognitive Sciences, 11(3), 105–110. 10.1016/j.tics.2006.12.002 [PubMed: 17254833]

Pickering MJ, & Garrod S. (2013). An integrated theory of language production and comprehension. The Behavioral and Brain Sciences, 36(4), 329–347. 10.1017/S0140525X12001495 [PubMed: 23789620]

Reuter T, Borovsky A, & Lew-Williams C. (2019). Predict and redirect: Prediction errors support children's word learning. Developmental Psychology, 55(8), 1656–1665. 10.1037/dev0000754 [PubMed: 31094555]

Sabourin L, Stowe LA, & De Haan GJ (2006). Transfer effects in learning a second language grammatical gender system. Second Language Research, 22(1), 1–29. 10.1191/0267658306sr259oa

Sheya A, & Smith L. (2018). Development weaves brains, bodies and environments into cognition. Language, Cognition and Neuroscience, 34(10), 1266–1273. 10.1080/23273798.2018.1489065

Slevc LR, & Novick JM (2013). Memory and cognitive control in an integrated theory of language processing. Behavioral and Brain Sciences, 36(4), 373–374.

Snedeker J, & Trueswell JC (2004). The developing constraints on parsing decisions: The role of lexical-biases and referential scenes in child and adult sentence processing. Cognitive Psychology, 49, 238–299. 10.1016/j.cogpsych.2004. [PubMed: 15342261]

Staub A, Abbott M, & Bogartz RS (2012). Linguistically guided anticipatory eye movements in scene viewing. Visual Cognition, 20(8), 922–946. 10.1080/13506285.2012.715599

Tanz C. (1980). Studies in the Acquisition of Deictic Terms. Cambridge: Cambridge University Press.

Trude AM (2013). When to simulate and when to associate? Accounting for inter-talker variability in the speech signal. Behavioral and Brain Sciences, 36, 375–376. 10.1017/S0140525X12002701

Trueswell JC, Sekerina I, Hill NM, & Logrip ML, (1999). The kindergarten-path effect: Studying on-line sentence processing in young children. Cognition, 73, 89–134. 10.1016/s0010-0277(99)00032-3 [PubMed: 10580160]

van Bergen G, & Flecken M. (2017). Putting things in new places: Linguistic experience modulates the predictive power of placement verb semantics. Journal of Memory and Language, 92, 26–42. 10.1016/j.jml.2016.05.003

Von Holzen K, & Mani N. (2012). Language nonselective lexical access in bilingual toddlers. Journal of Experimental Child Psychology, 113(4), 569–586. 10.1016/j.jecp.2012.08.001 [PubMed: 22980955]

Waxman SR (1999). Specifying the scope of 13-month-olds' expectations for novel words. Cognition, 70(3). 10.1016/S0010-0277(99)00017-7

Waxman SR, Lidz JL, Braun IE, & Lavin T. (2009). Twenty four-month-old infants' interpretations of novel verbs and nouns in dynamic scenes. Cognitive Psychology, 59(1), 67–95. 10.1016/j.cogpsych.2009.02.001 [PubMed: 19303591]

Wellman BH, Cross D, & Watson J. (2010). Meta-analysis of theory of mind development: The truth about false belief. Child Development 72. Child Development, 72(3), 655–684. 10.1111/1467-8624.00304

Wittenberg E, Khan M, & Snedeker J. (2017). Investigating thematic roles through implicit learning: Evidence from light verb constructions. Frontiers in Psychology, 8(JUN), 1–8. 10.3389/fpsyg.2017.01089 [PubMed: 28197108]

Ylinen S, Bosseler A, Junttila K, & Huotilainen M. (2016). Predictive coding accelerates word recognition and learning in the early stages of language development. Developmental Science, 1–13. 10.1111/desc.12472

Yoon S. & Brown-Schmidt S. (2013). What is the context of prediction? Behavioral and Brain Sciences, 36, 48–49.

Yurovsky D, Case S, & Frank MC (2017). Preschoolers flexibly adapt to linguistic input in a noisy channel. Psychological Science, 28(1), 132–140. 10.1177/0956797616668557 [PubMed: 28078978]

Yurovsky D, Smith LB, & Yu C. (2012). Does Statistical Word Learning Scale? It's a Matter of Perspective. In Miyake N, Peebles D, & Cooper R. (Eds.), Proceedings of the 34nd Annual Conference of the Cognitive Science Society. Austin, TX: Cognitive Science Society.
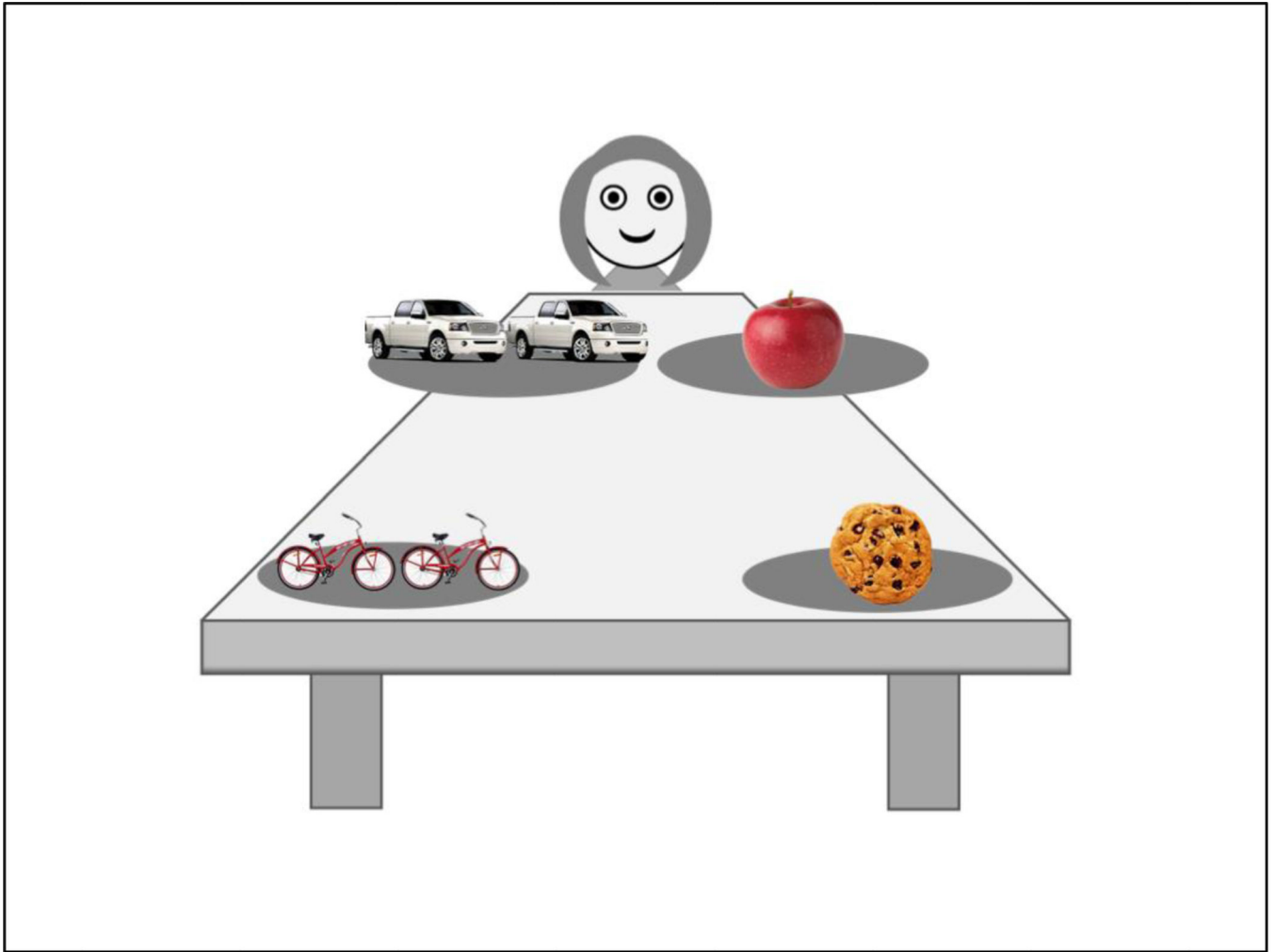
Figure 1:
Sample test trial for Experiment 1. During each test trial, participants heard a sentence referring to one of the four images (e.g., *Look at the/that wonderful cookie*).
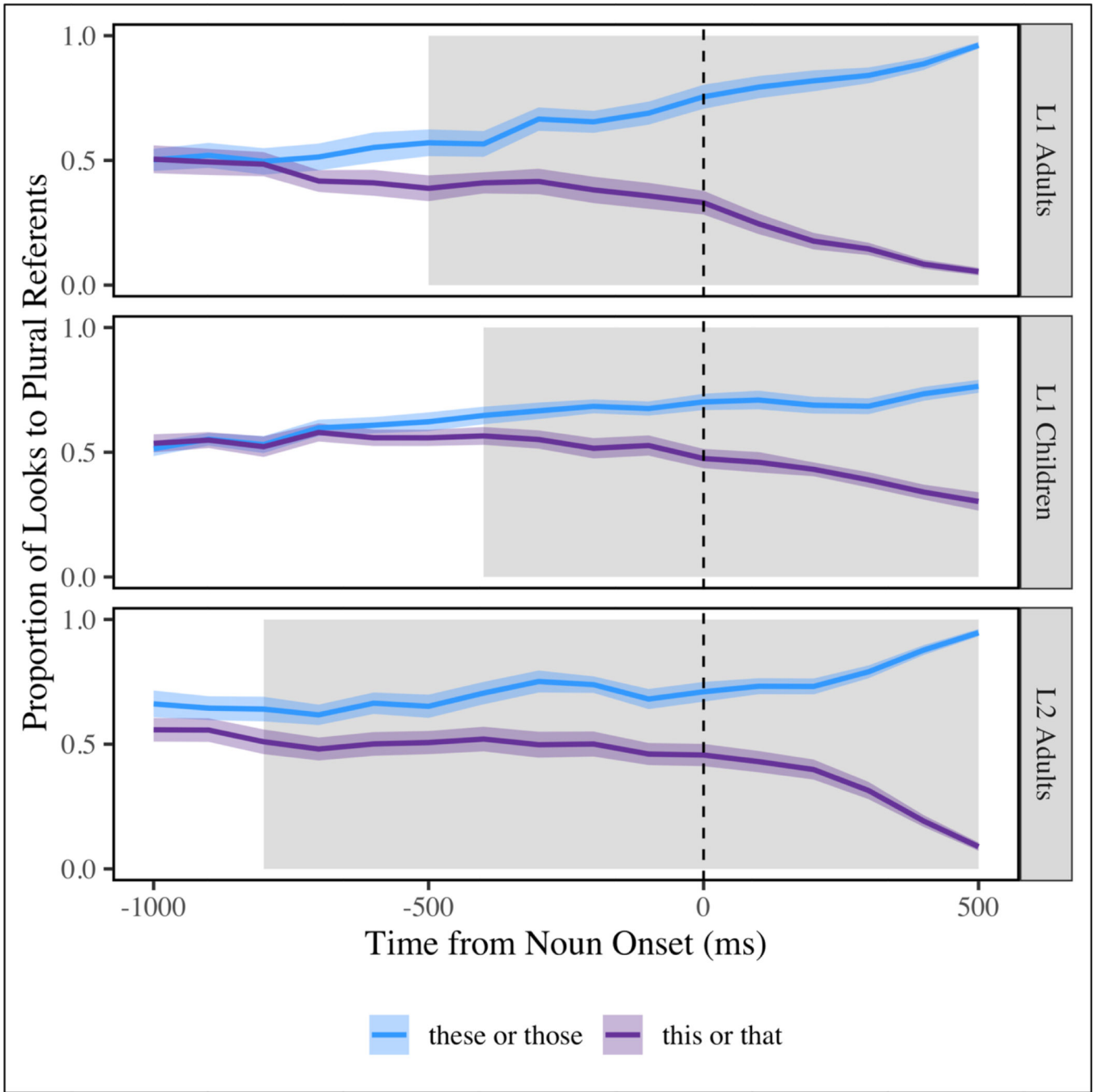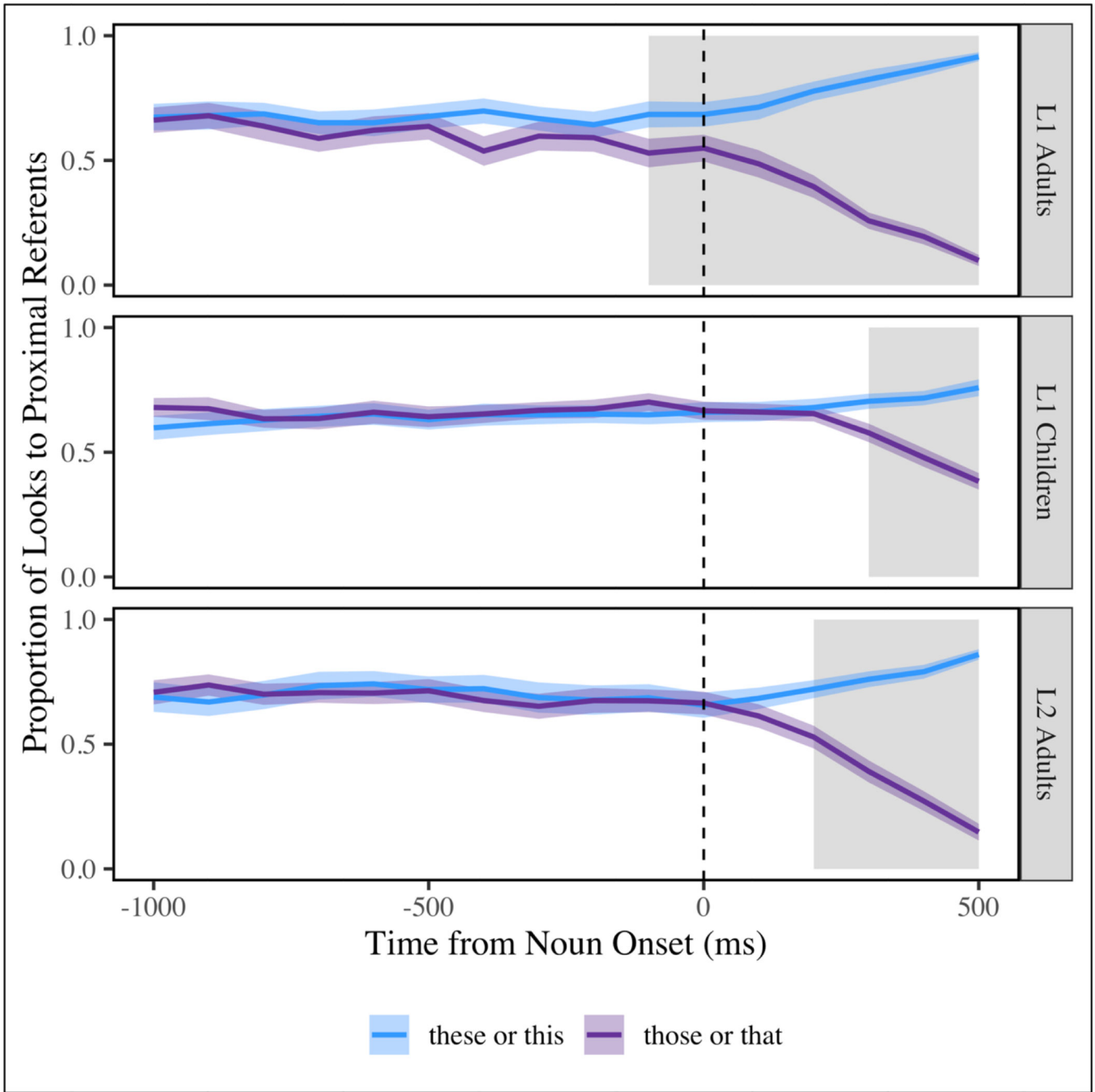
Figure 2:
Results from Experiment 1. Proportion of looks to plural referents for L1 adults ($n =$ 28), L1 children ($n = 28$), and L2 adults ($n = 28$) during plural deictic sentences (blue) and singular deictic sentences (purple). Line shading represents one standard error from the mean, averaged by subjects. Vertical dashed lines indicate noun onset. Area shading indicates significant effects from cluster-based permutation analyses ($ps < 0.05$). Results indicate that L1 adults, L1 children, and L2 adults used the plurality of deictic determiners to

predict the plurality of the upcoming referent, as evidenced by anticipatory eye movements generated before the onset of the number-marked noun.

Figure 3:

Results from Experiment 1. Proportion of looks to proximal referents for L1 adults ($n = 28$), L1 children ($n = 28$), and L2 adults ($n = 28$) during proximal deictic sentences (blue) and distal deictic sentences (purple). Line shading represents one standard error from the mean, averaged by subjects. Vertical dashed lines indicate noun onset. Area shading indicates significant effects from cluster-based permutation analyses ($ps < 0.05$). Results indicate that only L1 adults used the proximity information encoded in deictic determiners to predict the proximity of the upcoming referent.
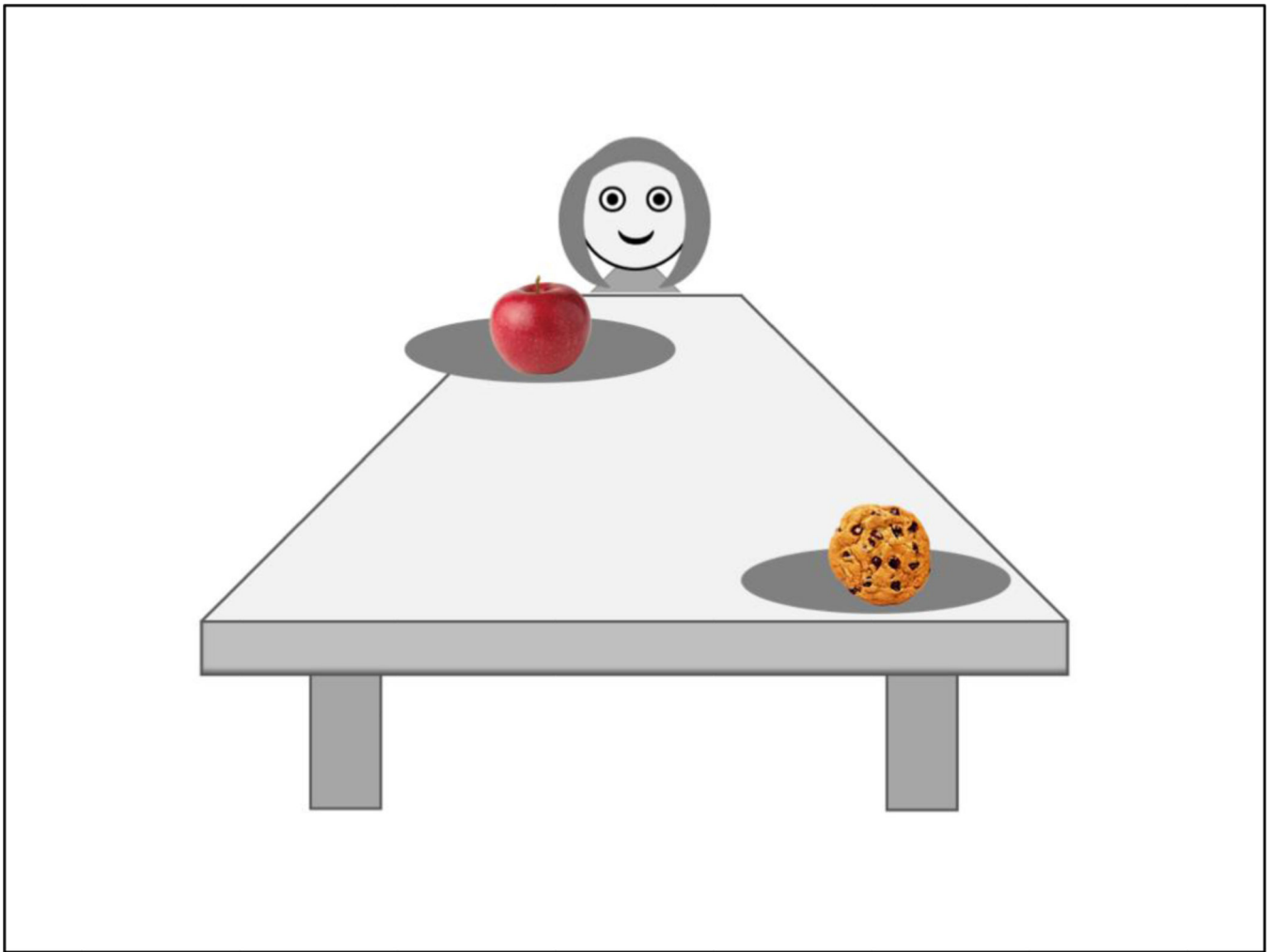
Figure 4:

Sample test trial for Experiment 2. The two visual referents were always located diagonally from each other. During each test trial, participants heard a sentence referring to one of the two images (e.g., *Look at the/that wonderful cookie*).
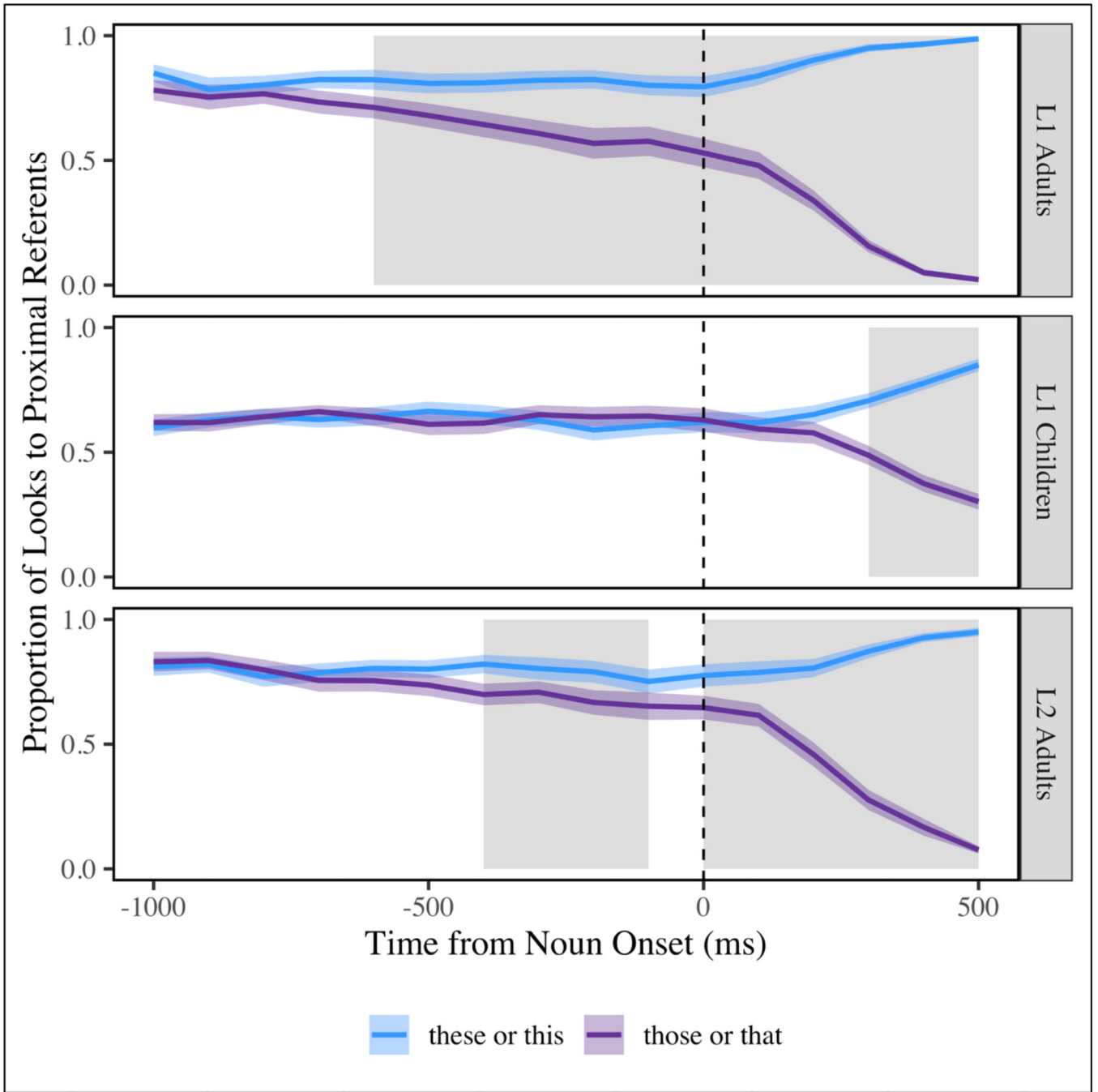
Figure 5:

Results from Experiment 2. Proportion of looks to proximal referents for L1 adults ($n = 28$), L1 children ($n = 28$), and L2 adults ($n = 28$) during proximal deictic sentences (blue) and distal deictic sentences (purple). Line shading represents one standard error from the mean, averaged by subjects. Vertical dashed lines indicate noun onset. Area shading indicates significant effects from cluster-based permutation analyses ($ps < 0.05$). Results indicate that L1 adults and L2 adults, but not L1 children, used deictic determiners to predict the

proximity of the upcoming referent, as evidenced by anticipatory eye movements generated before the onset of the proximal or distal noun.

Table 1

Experiment 1 descriptive statistics for L2 adults' self-reported English proficiency measures

Self-Report Measure	Min	Max	Mean	SD
Proficiency in Speaking English	4	9	7.57	1.50
Proficiency in Understanding English	5	9	7.96	1.20
Proficiency in Reading English	6	9	8.21	0.96
Proficiency in Writing English	4	9	7.71	1.46
Accent when Speaking English	1	9	6.39	2.38
Comfort when Speaking English	3	9	7.25	1.80

Table 2

Experiment 2 descriptive statistics for L2 adults' self-reported English proficiency measures

Self-Report Measure	Min	Max	Mean	SD
Proficiency in Speaking English	3	9	7.96	1.32
Proficiency in Understanding English	4	9	8.39	1.13
Proficiency in Reading English	5	9	8.46	0.92
Proficiency in Writing English	3	9	7.86	1.46
Accent when Speaking English	3	9	6.93	1.54
Comfort when Speaking English	5	9	7.86	1.15