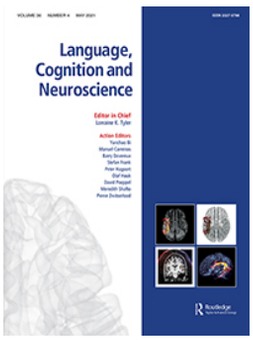

Language, Cognition and Neuroscience

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/plcp21

Adults and children predict in complex and variable referential contexts

Tracy Reuter, Kavindya Dalawella & Casey Lew-Williams

To cite this article: Tracy Reuter, Kavindya Dalawella & Casey Lew-Williams (2021) Adults and children predict in complex and variable referential contexts, Language, Cognition and Neuroscience, 36:4, 474-490, DOI: 10.1080/23273798.2020.1839665

To link to this article: https://doi.org/10.1080/23273798.2020.1839665

LANGUAGE, COGNITION AND NEUROSCIENCE
2021, VOL. 36, NO. 4, 474–490
https://doi.org/10.1080/23273798.2020.1839665

Routledge
Taylor & Francis Group

REGULAR ARTICLE

Adults and children predict in complex and variable referential contexts

Tracy Reuter, Kavindya Dalawella and Casey Lew-Williams

Department of Psychology, Princeton University, Princeton, NJ, USA

ABSTRACT

Prior research suggests that prediction supports language processing and learning. However, the ecological validity of such findings is unclear because experiments usually include constrained stimuli. While theoretically suggestive, previous conclusions will be largely irrelevant if listeners cannot generate predictions in response to complex and variable perceptual input. Taking a step toward addressing this limitation, three eye-tracking experiments evaluated how adults ($N = 72$) and 4- and 5-year-old children ($N = 72$) generated predictions in contexts with complex visual stimuli (Experiment 1), variable speech stimuli (Experiment 2), and both concurrently (Experiment 3). Results indicated that listeners generated predictions in contexts with complex visual stimuli *or* variable speech stimuli. When both were more naturalistic, listeners used informative verbs to generate predictions, but not adjectives or number markings. This investigation provides a test for theories claiming that prediction is a central learning mechanism, and calls for further evaluations of prediction in naturalistic settings.

ARTICLE HISTORY
Received 7 May 2020
Accepted 6 October 2020

KEYWORDS
Prediction; ecological validity; language processing; language development; anticipatory eye movements

A central tenet of prediction-based theories of language development is that learners generate predictions over a multitude of day-to-day processing experiences (Chater et al., 2016; Dell & Chang, 2014; Elman, 1990, 2009). According to this developmental viewpoint, "acquiring language is no more than acquiring the ability to process language" (Chater et al., 2016). By anticipating upcoming information, learners can use the same general computation (i.e., prediction) at different levels of representation (e.g., phonology, syntax, semantics) to incrementally acquire the correct forms and patterns of their native language(s). Importantly, real-world language processing contexts are inherently complex and variable. Learners must process and learn from variable auditory input, such as speakers with unfamiliar accents, and complex visual input, such as cluttered referential contexts. Although the inherent complexity and variability of naturalistic input creates many challenges for language processing and learning, prediction may give learners a powerful means of surmounting these challenges and may ultimately give rise to the speed and accuracy of adult language processing.

However, there is a disconnect between these theoretical claims and current empirical evidence: Experiments evaluating prediction have typically relied on simplified, repetitive referential contexts, and the ecological validity of such findings is therefore unclear (Huettig, 2015; Huettig & Mani, 2016). For example, eye-tracking experiments frequently use the visual world paradigm (Eberhard et al., 1995; Tanenhaus et al., 1995; for review see Kamide, 2008) or the looking-while-listening paradigm (Fernald et al., 2008; Lew-Williams & Fernald, 2009) to evaluate how listeners process visual and auditory information from moment to moment. In these paradigms, listeners view a small number of visual referents paired with isolated sentences which tend to repeat the same informative cues from trial to trial, such as verbs, adjectives, and number markings, with interspersed neutral trials. For instance, Altmann and Kamide (1999) tracked adults' eye movements as they viewed four referents and heard sentences with informative or neutral verbs (e.g., *The boy will eat the cake* vs. *The boy will move the cake*). Listeners' anticipatory eye movements suggest they can exploit informative verbs (e.g., *eat*) to predict the upcoming noun (e.g., *cake*). Findings from these carefully constrained paradigms provide an important step toward understanding human language processing, but results must be interpreted with caution. The ecological validity of empirical findings is largely unknown because prior research has relied extensively on what may be considered "prediction-encouraging

CONTACT Tracy Reuter treuter@alumni.princeton.edu; treuter@princeton.edu Department of Psychology, Princeton University, Princeton, NJ, 08544, USA

Supplemental data for this article can be accessed https://doi.org/10.1080/23273798.2020.1839665

experimental set-ups" (Huettig & Mani, 2016) that only demonstrate "what listeners *can* do, not what they actually do" (Huettig et al., 2011).

There are plausible reasons to suspect that prediction may be less likely *or* more likely to occur in naturalistic settings, as compared to more constrained lab contexts. On the one hand, the visual complexity and auditory variability of naturalistic contexts may pose challenges for prediction. For instance, if many visual referents are competing for attention, it may be difficult for listeners to rapidly and accurately identify the speaker's intended referent prior to explicit labelling.[1] It may also be more challenging to keep pace with a speaker's utterances if they contain semantic and syntactic variability across time, as compared to the more consistent auditory stimuli typically used in lab experiments. Such complexity and variability could scatter a listener's attentional focus and/or affect working memory, which has been linked to successful prediction (Federmeier, 2007; Ito et al., 2018; Otten & Van Berkum, 2009; Pickering & Gambi, 2018). However, on the other hand, the complexity and variability of naturalistic referential contexts may actually increase listeners' propensity to predict, chiefly because naturalistic contexts may be more engaging via a rich suite of visual, linguistic, and paralinguistic cues. For example, naturalistic contexts contain diverse arrays of objects (Luck, 2012), variable prosodic contours (Barthel et al., 2017; Nencheva et al., 2020), and speech disfluencies (Kidd et al., 2011). When bottom-up perceptual input is "noisy" or ambiguous, listeners may be more likely to engage top-down predictive mechanisms to compensate (Pickering & Gambi, 2018; Pickering & Garrod, 2007). The extent to which prediction underlies naturalistic conversations remains to be determined, and prediction may be more likely or less likely to occur in everyday referential contexts, as compared to more constrained laboratory environments.

To further explore whether and how prediction occurs during day-to-day language processing, a growing number of investigations have evaluated how adult listeners generate predictions within more naturalistic referential contexts (Andersson et al., 2011; Bögels, 2020; Coco et al., 2016; Sorensen & Bailey, 2007; Staub et al., 2012). For instance, an eye-tracking study by Staub et al. (2012) found that adults used informative verbs (e.g., *The woman will pour the coffee*) to predict a speaker's intended referent within a complex visual scene (e.g., a coffee pot on a cluttered kitchen counter). These findings suggest that adult listeners, when faced with the inherent disarray of real-world visual contexts, can rapidly and accurately process visual and auditory information and efficiently locate a relevant referent before it is named. Similarly, recent findings suggest that adults use prediction to successfully navigate "noisy" or variable auditory input (Gibson et al., 2013; Yurovsky et al., 2017; but see Brouwer et al., 2013). For example, Yurovsky et al. (2017) found that adult listeners used the plausibility of a speaker's prior utterances as a basis for interpreting noisy, perceptually ambiguous sentences (e.g., *I had carrots and peas/bees for dinner*). These results suggest that adult listeners are capable of generating predictions in the face of variable auditory information, using prior knowledge to infer the speaker's intended meaning. Generally, research showing accurate predictions despite complex and variable perceptual input lends preliminary support to the notion that prediction facilitates adults' language processing and learning over the course of natural conversational experiences.

Although investigations with adult listeners have made progress towards assessing prediction within more naturalistic referential circumstances, to our knowledge, developmental studies evaluating prediction have overwhelmingly done so within highly constrained experimental contexts (Andreu et al., 2013; Bobb et al., 2016; Borovsky et al., 2012; Borovsky et al., 2013; Borovsky & Creel, 2014; Creel, 2012, 2014; Fernald et al., 2008; Gambi et al., 2018; Kidd et al., 2011; Lew-Williams, 2017; Lew-Williams & Fernald, 2007; Lukyanenko & Fisher, 2016; Mani & Huettig, 2012; Nation et al., 2003; Reuter et al., 2019). Some developmental studies have included multiple referents or multiple sentence constructions across trials, or at minimum have included a large number of neutral trials (e.g., Altmann & Kamide, 2007; Snedeker & Trueswell, 2004). Similarly, prior findings suggest that children may be able to generate and update predictions when faced with more variable auditory input (Havron et al., 2019; Yurovsky et al., 2017). However, in our view, developmental research has yet to systematically investigate how young listeners generate predictions in referential contexts with even semi-naturalistic visual complexity and auditory variation. Thus, while prediction occurs *in theory* during a multitude of learners' everyday language processing experiences, it is most frequently evaluated within constrained, repetitive experimental contexts. The case for prediction as a developmental mechanism hinges on children's ability to generate predictions within more naturalistic referential contexts.

Relying on constrained developmental paradigms has resulted in uncertain ecological validity, but it is also worth noting that these paradigms have enabled careful examination of how young listeners interpret and predict information during real-time processing. Many clever experimental designs have positioned developmental and language scientists to better

understand children's emergent prediction abilities. Findings from these paradigms suggest that children can use a variety of linguistic and paralinguistic cues to more rapidly recognise words that occur later in the sentence, including: lexical semantics (Fernald et al., 2008, 2010; Mani & Huettig, 2012), morphosyntax (Lew-Williams, 2017; Lew-Williams & Fernald, 2007; Lukyanenko & Fisher, 2016; Reuter et al., 2020), speech disfluencies (Kidd et al., 2011), and speaker identity (Borovsky & Creel, 2014; Creel, 2012, 2014). For example, Mani and Huettig (2012) found that 2-year-old children, while viewing two referents (e.g., a cake and a bird) could use informative verbs (e.g., *The boy eats the big cake*) to predict the upcoming edible referent. Similarly, Fernald et al. (2010) found that 3-year-old children could use adjectives (e.g., *Which one's the blue/red car?*) to more rapidly identify one of two available referents (e.g., a blue car and a red car). Recent findings also suggest that children can use morphosyntactic cues such as number markings as a basis for prediction. For example, while viewing plural and singular referents (e.g., two cookies and one apple), 3-year-old children can exploit informative number markings (e.g., *Where are the good cookies?* vs. *Where is the good apple?*) to accurately predict upcoming referents (Lukyanenko & Fisher, 2016). Similarly, Reuter et al. (2020) found that 4- and 5-year-old children could use deictic number markings (e.g., *Look at that nice cookie* vs. *Look at those nice cookies*) to rapidly anticipate upcoming singular and plural referents, respectively. These eye-tracking experiments, although unlikely to reflect the natural variability of real-world conversations, have laid a foundation of empirical evidence which suggests that children might be capable of generating predictions during real-time language processing.

In order to determine the relevance of prediction for language development, it is important to evaluate prediction not only in constrained processing contexts, but in contexts that include natural complexity and variability in visual and auditory information. In the end, such evaluations will need to take place in completely natural contexts that reflect the lived, dynamic communicative experiences of infants and young children – not only at particular ages, but across development. Here, we took an incremental step toward this ideal by evaluating how adults and 4- to 5-year-old children generate predictions in experimental contexts with somewhat complex visual stimuli and somewhat variable speech stimuli. In Experiment 1, we assessed how listeners generate predictions using informative verbs when visual stimuli are somewhat complex. In Experiment 2, we assessed how listeners generate predictions using informative verbs, adjectives, and number markings when visual stimuli are relatively simple. Finally, in Experiment 3, we manipulated both visual and auditory stimuli, assessing how listeners generate predictions using informative verbs, adjectives, and number markings when visual scenes are somewhat complex and auditory stimuli are somewhat variable. The three experiments allowed for comparisons to the empirical studies that inspired their designs (specifically: Fernald et al., 2008; Fernald et al., 2010; Lukyanenko & Fisher, 2016; [blind for peer review]), but extended current knowledge by evaluating prediction within incrementally more naturalistic processing contexts. In doing so, this work addresses the limited ecological validity of prior studies and evaluates theories claiming that prediction is a key mechanism supporting language development.

Experiment 1

In Experiment 1, we assessed how adults and children generate predictions in speech when visual stimuli are more complex than in previous lab studies. Rather than viewing isolated images, participants viewed photographic images of indoor scenes with multiple, overlapping visual referents that varied naturally in size, colour, and spatial location. Participants heard two types of sentences: Predictive sentences included semantically-informative verbs that listeners could use to predict the upcoming target (e.g., *Dan dealt the cards on the table*), whereas neutral sentences did not include any words that supported early identification of the target (e.g., *Dan noticed the cards on the table*). The design of Experiment 1 was similar to prior eye-tracking studies with adult participants (Coco et al., 2016; Staub et al., 2012). We conducted a near-replication of this study and, furthermore, extended it to include 4- to 5-year-old children.

Method

Participants

Experiment 1 participants were 24 monolingual, English-speaking adults (9 male) and 24 children (14 male) from monolingual, English-speaking households. Adults were 18–22 years old (M = 19.5 years, SD = 1.44 years) and children were 47–68 months old (M = 60 months, SD = 6.11 months). Adult participants were recruited from the Princeton University campus via course credit and paid study pools and families were recruited from nearby communities via research outreach events, fliers, and social media. Participants had no known hearing or vision impairments. We tested one additional child participant but excluded them from analyses due to experimenter error. The Princeton University Institutional

Review Board approved this research protocol (IRB record number 7117). Experimenters obtained informed consent from all adult participants and from a legal guardian of all child participants. Adult participants received credit for psychology courses or payment ($8 USD) as compensation, and families received payment ($10 USD), a children's book, and a children's t-shirt as compensation for their time.

Stimuli and design

Speech stimuli consisted of two types of pre-recorded sentences. Predictive sentences included semantically-informative verbs that listeners could use to predict an upcoming noun (e.g., *Dan dealt the cards on the table*), whereas neutral sentences did not include any words that could be used to predict the upcoming noun (e.g., *Dan noticed the cards on the table*). A female, native speaker of English recorded stimuli for Experiments 1, 2, and 3, using child-directed intonation. For each auditory stimulus, we used Praat (Boersma & Weenink, 2017) to measure the onset of the informative verb and the onset of the target noun. On average, the onset of the informative verb occurred 759 ms prior to the onset of the target noun for predictive sentences. For neutral sentences, the uninformative verb occurred 752 ms prior to target noun onset, on average. Predictive and neutral conditions did not vary significantly in time from verb onset to noun onset (paired-sample $t(11) = 0.11$, $p = 0.914$).

Visual stimuli were photographic images of six indoor scenes: a kitchen sink, a kitchen stove, a kitchen counter, a dining room, a home office, and a living room. Scenes contained multiple, varied household objects (Figure 1). Each image was 1280×1024 pixels. Photographing scenes from two angles (right and left) created two exemplars for each visual stimulus. Each visual stimulus appeared twice during the task. One exemplar appeared with a neutral sentence and the other exemplar appeared with a predictive sentence.

During each trial, visual stimuli appeared 4 s prior to the onset of speech stimuli and remained visible for 2 s, such that the total duration of each trial was 6 s. We determined preview times for each experiment based on the time listeners would presumably need to encode visual referents and their locations – an essential step for generating accurate anticipatory eye movements. The Experiment 1 preview time (4 s) is longer than those used in comparable studies with adult listeners (Coco et al., 2016; Staub et al., 2012). However, to our knowledge, no prior studies have evaluated *children's* prediction abilities with more complex visual stimuli. We therefore used a preview time of 4 s for Experiment

Figure 1. Example of a visual scene from Experiment 1. Participants viewed photographic stimuli and heard either a neutral sentence (e.g., *Dan dealt the cards on the table*) or a predictive sentence (e.g., *Dan noticed the cards on the table*).

1 to give adults and children enough time to encode visual referents prior to the onset of the auditory stimuli.

Trials appeared in one of four quasi-randomized orders, which counterbalanced visual stimulus exemplars across orders, and ensured that condition (neutral or predictive) did not repeat for more than three trials sequentially. Filler trials occurred every four trials, and consisted of a cartoon image (e.g., a smiling girl) and encouraging statements (e.g., "You're doing great! Keep it up!"). In total, Experiment 1 included 6 predictive trials, 6 neutral trials, and 3 filler trials. All visual and auditory stimuli and experiment code for Experiment 1 are available on the Open Science Framework.

Procedure

The study took place in a sound-attenuated room at the Princeton Baby Lab. Participants sat approximately 60 cm from an EyeLink 1000 Plus eye-tracker. Child participants sat in a booster seat. The experimenter controlled the study from a Mac host computer, using EyeLink Experiment Builder software (SR Research, Mississauga, Ontario, Canada). Before beginning the study, the experimenter first placed a target sticker on the participant's face to allow the eye-tracker to record their eye movements, and then calibrated the eye-tracker for each participant with a standard five-point calibration procedure. Throughout the task, participants viewed stimuli on a 17-inch LCD monitor and the eye tracker recorded their eye movements with a sampling rate of 500 Hz. The total duration of the study was five minutes.

Results

During the experiment, the eye tracker automatically recorded participants' fixations every 2 ms (500 Hz). We analysed samples recorded within a 500×500 pixel area surrounding each visual referent (which included some overlapping referents) and eliminated any samples that were outside of these visual areas of interest (406,564 of 2,631,022 samples, 16%) prior to aggregating data within 100-ms time-bins. In order to assess whether participants used informative verbs to predict the upcoming target nouns, we analysed participants' looking behaviour during a time window from 1000 ms before to 1000 ms after the onset of the target noun. If listeners predict the upcoming referent, then we expected to observe the emergence of condition effects *before* the onset of the target noun (0 ms), indicating that listeners generated anticipatory eye movements to the target.

We analysed listeners' proportion of target looks during neutral and predictive sentences with a mixed-effects logistic regression model, using the lme4 package (Version 1.1-21; Bates et al., 2015) and the lmerTest package (Version 3.1-0; Kuznetsova et al., 2017). The model included fixed effects for age group (adults, children), condition (neutral, predictive) and time (100-ms bins, −1000 to 1000 ms from noun onset) as well as their interactions. The model also included random intercepts for subjects and items, which was the maximal model that converged (Barr et al., 2013). Model results revealed significant effects for condition ($\beta = -0.48$, $z = -13.87$, $p < 0.001$), age group ($\beta = 0.33$, $z = 3.83$, $p < 0.001$), and time ($\beta = 1.24$, $z = 22.41$, $p < 0.001$), indicating that listeners' target looks were greater for predictive trials than for neutral trials, that adults generated more target looks than children, and that listeners' target looks increased over time. Model results also revealed an interaction of condition and time ($\beta = -0.17$, $z = -3.07$, $p = 0.002$) and an interaction of age group and time ($\beta = 0.34$, $z = 6.21$, $p < 0.001$), indicating that the condition difference (predictive > neutral) was greater at earlier time points than at later time points, and that adults' target looks increased at a faster rate than children's target looks over time. Finally, model results indicated a three-way interaction of condition, age group, and time ($\beta = 0.14$, $z = 2.53$, $p = 0.012$), indicating that adults' interaction effect for condition and time was more robust than that of children. Together, results suggest that both adults and children used informative verbs to anticipate upcoming referents.

We next analysed adults' and children's looking behaviours with cluster-based permutation analyses (Maris & Oostenveld, 2007), in order to match the analytical approaches of prior eye-tracking studies (Reuter et al., 2020; Wittenberg et al., 2017). Findings revealed significant clusters that emerged prior to the onset of the target noun for adults (−600 to 800 ms, cluster $t = 41.69$, $p < 0.001$) and for children (−300 to 1000 ms, cluster $t = 39.46$, $p < 0.001$). Together, results from mixed-effects regression and cluster-based permutation analyses suggest that adults and children used informative verbs to predict the identity of target referents, although adults were more proficient in rapidly and accurately orienting to the target referent (Figure 2).

Discussion

Experiment 1 findings suggest that both adult and child listeners can take advantage of semantically-informative verbs during real-time language processing within complex visual scenes to rapidly and accurately identify upcoming referents. Upon hearing a semantically-informative verb (e.g., *dealt*), adults and children efficiently launched anticipatory eye movements to the

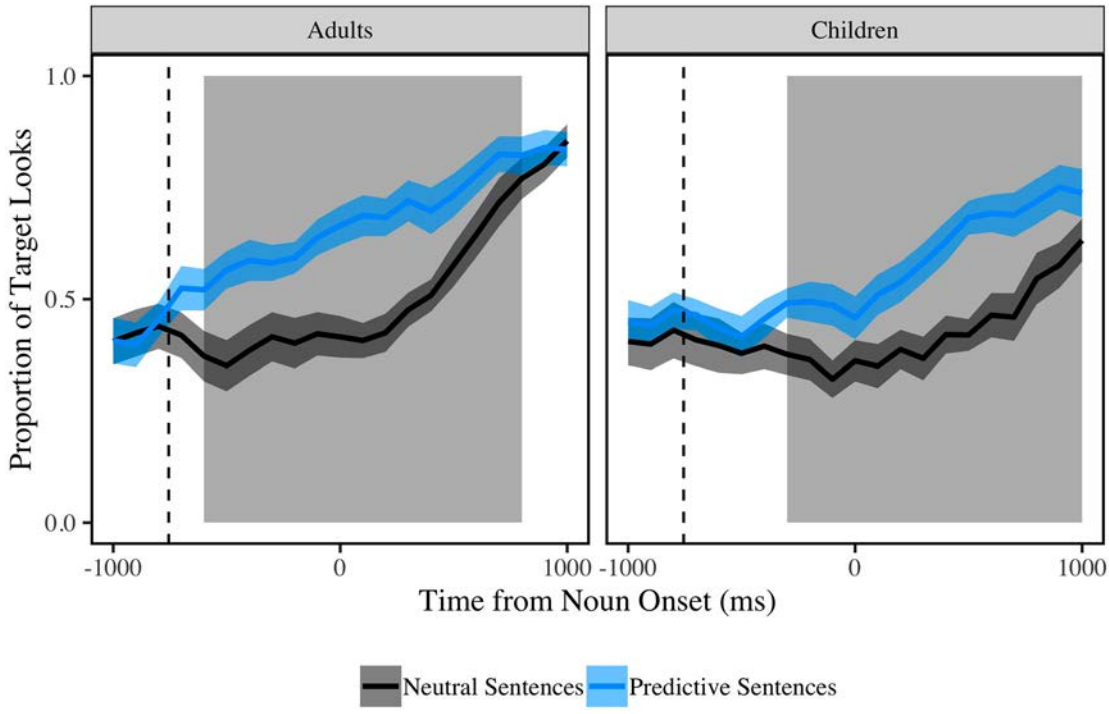

Figure 2. Looking-time plots for Experiment 1. Proportion of looks to the target referent during neutral sentences (grey) and predictive sentences (blue) are shown for adults ($n = 24$) and for children ($n = 24$). The onset of the target noun (e.g., *cards*) is at 0 ms. Vertical dashed lines indicate the average onset of the informative verb in predictive sentences (e.g., *dealt*). Line shading indicates one standard error from the mean for each condition, averaged by subjects. Area shading indicates significant clusters ($ps < 0.05$) from permutation analyses (Maris & Oostenveld, 2007). Results indicate that both adults and children generated anticipatory eye movements to the target referent during predictive sentences, and further suggest that listeners can use informative verbs to pre-activate upcoming representations during language processing.

appropriate visual referent (e.g., *cards*). This pattern of results provides a conceptual replication of prior research with adult participants (Coco et al., 2016; Staub et al., 2012) and further suggests that children, much like adults, can generate predictions (specifically, verb-based predictions) in somewhat varied, naturalistic language processing contexts. Although the observed prediction effect was more robust for adults, suggesting that further development may be necessary for children to eventually match the speed and accuracy of adults' predictions, these developmental findings suggest that prediction may be a learning mechanism that operates in at least somewhat ecologically sound processing contexts.

However, although Experiment 1 visual stimuli were more naturalistic than in prior investigations, the speech stimuli were highly constrained. Every trial included a verb (e.g., *dealt*, *noticed*) and the verb was informative for identifying the target referent on 50% of the trials. This experimental design is problematic for two principle reasons. First, the parallel structure of sentences from trial to trial may have guided listeners (with or without their explicit awareness) to attend to the verbs. Rather than strictly using semantically-informative verbs to generate predictions, it is possible that listeners were able to make inferences about the structure of the task. A second, broader limitation of this design is that the repetitive structure may not reflect real-world language processing circumstances, as listeners must typically contend with diverse speech across time in naturalistic processing contexts. Thus, in Experiment 2, we manipulated the variability of the speech stimuli, but used relatively simple visual stimuli.

Experiment 2

In Experiment 2, we evaluated how adults and children generate predictions in processing contexts that involve more variation in speech stimuli across the course of the experiment. Participants viewed constrained visual stimuli with two side-by-side, colour-matched referents (one single and one plural), as in many prior studies, and they heard a combination of predictive sentences and neutral sentences. Importantly, unlike Experiment 1, the informative cue for predictive sentences varied from trial to trial, including informative verbs (e.g., *Do you want to read the yellow books?*), size adjectives (e.g., *Do you see the little yellow books?*), and

number markings (e.g., *Where are the yellow books?*; *Can you see those yellow books?*). Neutral sentences were also varied (e.g., *Do you see the yellow books?*; *Can you find the yellow books?*; *Can you show me the yellow books?*; and *Can you see the yellow books?*). Prior developmental findings suggest that children can use these cues, among others, as a basis for generating predictions during real-time language processing (Fernald et al., 2008, 2010; Lew-Williams, 2017; Lukyanenko & Fisher, 2016; Mani & Huettig, 2012; Reuter et al., 2020).

Method

Participants

Experiment 2 participants were 24 monolingual, English-speaking adults (7 male) and 24 children (15 male) from monolingual, English-speaking households. Adults were 18–27 years old ($M = 19.79$ years, $SD = 2.04$ years) and children were 49–71 months old ($M = 60.25$ months, $SD = 5.89$ months). Participants were recruited via the same routes as those in Experiment 1. Participants had no known hearing or vision impairments. We tested one additional child participant but excluded them from analyses due to inattention during the study. The Princeton University Institutional Review Board approved this research protocol (IRB record number 7117). Experimenters obtained informed consent from all adult participants and from a legal guardian of all child participants. Participants received the same compensation for their time as those in Experiment 1.

Stimuli, design, and procedure

Auditory stimuli included two types of pre-recorded sentences. Predictive sentences included informative cues – verbs (Mani & Huettig, 2012), adjectives (Fernald et al., 2010), number markings (Lukyanenko & Fisher, 2016), and deictic number markings (Reuter et al., 2020) – that listeners could use to predict an upcoming noun (e.g., verbs: *Do you want to read the yellow books?*; size adjectives: *Do you see the little yellow books?*; number markings: *Where are the yellow books?*; and deictic number markings: *Can you see those yellow books?*). Neutral sentences were likewise varied, but did not include any words that could be used to predict the upcoming noun (e.g., *Do you see the yellow books?*; *Can you find the yellow books?*; *Can you show me the yellow books?*). The average onset of the informative cue was 1210 ms before the onset of the target noun for predictive sentences. For neutral sentences, the uninformative verb occurred 1615 ms prior to target noun onset, on average. Neutral sentences had a greater time duration between the onset of uninformative verbs (e.g., *see*) and target nouns, as compared to the duration between the onset of the informative cues (e.g., *little*) and target nouns in predictive sentences (paired-sample $t(63) = 13.51$, $p < 0.001$).

Visual stimuli were singular and plural images of the eight target nouns: *ball, bike, book, cat, chair, cookie, slide,* and *truck*. Each target image was approximately 450×450 pixels and appeared on a 500×500 pixel white background. Visual stimuli appeared in yoked pairs (i.e., *ball-cat, bike-truck, book-chair*, and *cookie-slide*). Each yoked pair appeared eight times during the experiment (four times with a neutral sentence and four times with a predictive sentence; four times with a singular target and four times with a plural target). Importantly, the images for each yoked pair were matched in colour (Figure 3). This matching process ensured that the colour adjective included in each speech stimulus did not provide information that could be used to identify the target image. Rather, the inclusion of the colour adjective gave listeners additional time to generate predictions and to launch anticipatory eye movements to the target image.

During each trial, visual stimuli appeared for 500 ms prior to the onset of speech stimuli and remained visible for 3 s, such that the total duration of each trial was 3.5 s. The preview time for Experiment 2 (500 ms) followed the same rationale as for Experiment 1: We expected that adults and children would be able to rapidly encode two simple visual referents. Trials appeared in one of eight quasi-randomized orders, which counterbalanced target plurality (singular or plural), target side (right or left), and ensured that condition (neutral or predictive) and visual stimulus yoked pair (*ball-cat, bike-truck, book-chair*, and *cookie-slide*) did not repeat for more than three trials sequentially. Filler trials occurred every eight trials, and consisted of a cartoon image (e.g., a smiling boy) and encouraging statements (e.g., "Great work! Let's try some more"). In total, Experiment 2 included 16 predictive trials, 16 neutral trials, and 4 filler trials. Other procedural details for Experiment 2 were identical to those of Experiment 1. All visual and auditory stimuli and experiment code for Experiment 2 are available on the Open Science Framework.

Results

As in Experiment 1, we analysed samples recorded within a 500×500 pixel area surrounding each visual referent and eliminated any samples that were outside of these visual areas of interest (620,309 of 4,327,280 samples, 14%) prior to aggregating data within 100-ms time-bins. We analysed participants' looking behaviour during a time window from 1000 ms before to

Figure 3. Example of visual stimuli in Experiment 2. Participants viewed stimuli with two colour-matched referents (one plural, one singular) and heard varied neutral sentences (e.g., *Do you see the yellow books?*) and varied predictive sentences that included informative verbs, size adjectives, or number markings (e.g., *Do you want to read the yellow books?*; *Do you see the little yellow books?*; *Where are the yellow books?*; and *Can you see those yellow books?*).

1000 ms after the onset of the target noun. If listeners can use informative verbs, size adjectives, and number markings in intermixed orders across trials to predict upcoming referents, then we expected to observe the emergence of condition effects *before* the onset of the target noun (0 ms).

We analysed listeners' proportion of target looks during neutral and predictive sentences with a mixed-effects logistic regression model with the same specifications as in Experiment 1. Model results revealed significant effects for condition ($\beta = -0.40$, $z = -21.69$, $p < 0.001$), age group ($\beta = 0.25$, $z = 4.36$, $p < 0.001$), and time ($\beta = 1.77$, $z = 57.96$, $p < 0.001$), indicating that listeners' target looks were greater for predictive trials than for neutral trials, that adults' target looks were greater than children's, and that listeners' target looks increased over time. Model results also revealed an interaction of condition and time ($\beta = 0.26$, $z = 8.46$, $p < 0.001$) and an interaction of age group and time ($\beta = 0.45$, $z = 14.74$, $p < 0.001$), indicating that the condition difference (predictive > neutral) increased over time, and that adults' target looks increased at a faster rate than children's. The three-way interaction of condition, age group, and time was not significant ($\beta = 0.01$, $z = 0.37$, $p = 0.709$). Together, Experiment 2 results suggest that both adults and children used diverse linguistic cues in across trials to anticipate upcoming referents.

As in Experiment 1, we next analysed adults' and children's looking behaviours with cluster-based permutation analyses using the same analytical steps as described previously. Findings revealed significant clusters which emerged prior to the onset of the target noun for adults (−1000 to 600 ms, cluster $t = 77.15$, $p < 0.001$) and for children (−1000 to 500 ms, cluster $t = 54.30$, $p < 0.001$). In sum, results from mixed-effects regression and cluster-based permutation analyses converged to suggest that both adults and children were able to exploit informative verbs, size adjectives, and number marking to predict the upcoming target nouns (Figure 4).

We next determined whether listeners used each type of predictive cue independently to anticipate upcoming information. To evaluate whether participants used informative verbs (e.g., *Do you want to read the yellow books?*), size adjectives (e.g., *Do you see the little yellow books?*), number markings (e.g., *Where are the yellow books?*) and deictic number markings (e.g., *Can you see those yellow books?*) to predict the upcoming target noun, we compared participants' proportion of target looks during each of the four types of predictive sentences with their proportion of target looks during neutral sentences. We analysed adults' and children's looking behaviours with mixed-effects logistic regression models, including interacting fixed effects for condition (neutral, predictive) and time (100-ms bins, −1000 ms to 0 ms from noun onset), and random intercepts for subjects and items, which was the maximal model that converged. Results, summarised in Table 1, indicate that both adults and children were capable of exploiting each type of informative word to anticipate upcoming information during real-time language processing. Cluster-based permutation analyses further revealed significant clusters which emerged prior to the onset of the target noun for

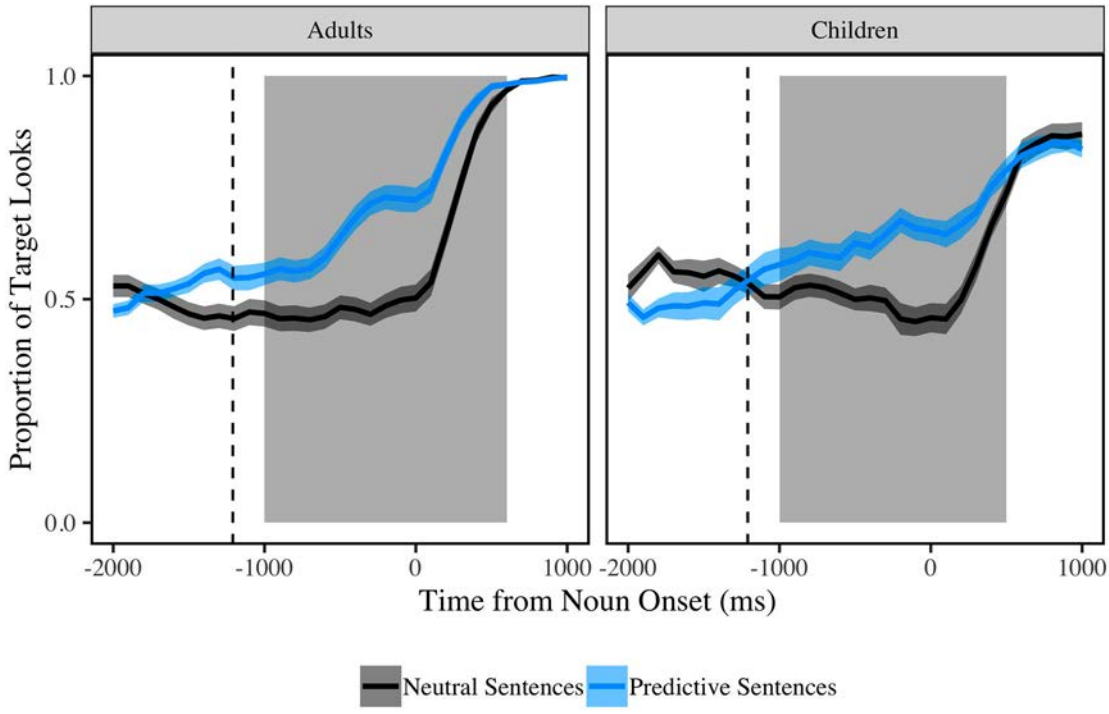

Figure 4. Looking-time plots for Experiment 2. Proportion of looks to the target referent during neutral sentences (grey) and predictive sentences (blue) for adults ($n = 24$) and for children ($n = 24$). The onset of the target noun (e.g., *books*) is at 0 ms. Vertical dashed lines indicate the average onset of the informative word in predictive sentences (e.g., *read*). Line shading indicates one standard error from the mean for each condition, averaged by subjects. Area shading indicates significant clusters ($ps < 0.05$) from permutation analyses (Maris & Oostenveld, 2007). Results indicate that adults and children generated anticipatory eye movements to the target referent during predictive sentences, further suggesting that listeners can flexibly use a variety of currently-available information – verbs, size adjectives, and number marking – to pre-activate upcoming representations during language processing. Post-hoc analyses including a split-halves analysis, a first occurrence analysis, and a trial-by-trial analysis are available in Supplementary Materials on the Open Science Framework.

adults (verbs: −700 to 600 ms, cluster $t = 50.18$, $p < 0.001$; size adjectives: −1000 to 500 ms, cluster $t = 49.58$, $p < 0.001$; number markings: −1000 to 600 ms, cluster $t = 48.21$, $p < 0.001$; deictic number markings: −300 to 500 ms, cluster $t = 27.99$, $p < 0.001$) and for children (verbs: −1000 to 400 ms, cluster $t = 51.25$, $p < 0.001$; size adjectives: −200 to 300 ms, cluster $t = 17.97$, $p < 0.001$; number markings: −300 to 600 ms, cluster $t = 22.67$, $p < 0.001$; deictic number markings: −800 to −600 ms, cluster $t = 4.74$, $p = 0.034$ and −200 to −400 ms, cluster $t = 16.34$, $p < 0.001$). Figures for each predictive cue are available in Supplementary Materials on the Open Science Framework.

Discussion

Findings from Experiment 2 suggest that both adult and child listeners can use informative verbs, size adjectives,

Table 1. Fixed effects for condition from mixed-effects logistic regression models for Experiments 2 and 3.

	Age group	Verbs	Size adjectives	Number markings	Deictic number markings
Experiment 2	Adults	$\beta = 2.36$ $t = 10.59$ ***	$\beta = 2.10$ $t = 9.54$ ***	$\beta = 1.51$ $t = 6.75$ ***	$\beta = 1.55$ $t = 6.96$ ***
	Children	$\beta = 2.25$ $t = 9.21$ ***	$\beta = 1.26$ $t = 4.94$ ***	$\beta = 1.23$ $t = 4.93$ ***	$\beta = 1.12$ $t = 4.14$ ***
Experiment 3	Adults	$\beta = 2.03$ $t = 11.52$ ***	$\beta = 0.71$ $t = 4.11$ ***	$\beta = 0.07$ $t = 0.41$	$\beta = 0.31$ $t = 1.80$ ~
	Children	$\beta = 2.24$ $t = 10.56$ ***	$\beta = 0.39$ $t = 1.86$ ~	$\beta = -0.28$ $t = -1.40$	$\beta = 0.25$ $t = 1.17$

Note: Significant effects (predictive > neutral) indicate that adults and children used available information (verbs, size adjectives, number marking, and deictic number marking) to predict the upcoming target noun (~$p < 0.10$, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$).

and number markings – paired with highly constrained visual stimuli – to efficiently identify upcoming referents. Upon hearing a semantically-informative verb (e.g., *read*), size adjective (i.e., *big/little*), or number marking (i.e., *is/are*; *that/those*), adults and children rapidly launched anticipatory eye movements to the target referent (e.g., *chair/books*). This pattern of results converges with prior developmental research indicating that children can use each of these cues as a basis for generating predictions (Fernald et al., 2008, 2010; Lew-Williams, 2017; Lukyanenko & Fisher, 2016; Mani & Huettig, 2012). By evaluating adults' and children's predictions in contexts with more varied auditory stimuli, these findings, combined with those of Experiment 1, begin to address the unknown ecological validity of prior results and suggest that prediction may be deployed in more naturalistic audiovisual environments, and therefore may be a plausible mechanism for language learning in childhood.

However, further work is needed to evaluate whether and how prediction occurs in more naturalistic language processing circumstances. Experiments 1 and 2 each manipulated one aspect of the communicative context while keeping the other aspect constrained, such that either the visual *or* auditory stimuli were more complex or variable relative to prior developmental investigations (e.g., Mani & Huettig, 2012). It therefore remains uncertain whether or not listeners can contend with *simultaneous* complexity and variation in visual and auditory stimuli. To further explore listeners' prediction abilities in more naturalistic contexts, we manipulated the complexity and variability of visual and auditory stimuli in Experiment 3.

Experiment 3

In Experiment 3, we aimed to assess how adults and children generate predictions in referential contexts with naturalistic visual complexity and variable linguistic cues across time. Participants viewed semi-constrained photographic stimuli which included four referents, and heard a combination of predictive sentences and neutral sentences. As in Experiment 1, the informative linguistic cue for predictive sentences varied from trial to trial, including informative verbs (e.g., *Could Sally eat the red apples?*), size adjectives (e.g., *Do you see the little red apples?*), and number markings (e.g., *Where are the red apples?*; *Can you see those red apples?*). Neutral sentences were also varied (e.g., *Do you see the red apples?*; *Can you find the red apples?*; *Can you show me the red apples?*; and *Can you see the red apples?*).

Method

Participants

Experiment 3 participants were 24 monolingual, English-speaking adults (8 male) and 24 children (9 male) from monolingual, English-speaking households. Adults were 18–23 years old ($M = 19.33$ years, $SD = 1.43$ years) and children were 48–71 months old ($M = 60.29$ months, $SD = 8.31$ months). Participants were recruited using the same procedures as those in Experiments 1 and 2. Participants had no known hearing or vision impairments. We tested three additional child participants but excluded them from analyses due to a previously-diagnosed developmental delay, inattention during the study, and experimenter error, respectively. The Princeton University Institutional Review Board approved this research protocol (IRB record number 7117). Experimenters obtained informed consent from all adult participants and from a legal guardian of all child participants. Participants received the same compensation for their time as those in Experiments 1 and 2.

Stimuli, design, and procedure

As in Experiment 2, auditory stimuli for Experiment 3 included varied predictive and neutral sentences. Predictive sentences included informative cues – verbs, size adjectives, and number marking – that listeners could use to predict an upcoming noun (e.g., *Could Sally eat the red apples?*; *Do you see the little red apples?*; *Where are the red apples?*; and *Can you see those red apples?*), whereas neutral sentences but did not include any words that could be used for prediction (e.g., *Do you see the red apples?*; *Can you find the red apples?*; *Can you show me the red apples?*). Each sentence contained a colour adjective that gave participants more time to process the main informative cue prior to noun onset; this colour adjective narrowed the number of potential referents from four to two, but did not reveal the identity of the target noun. The average onset of the informative cue was 1091 ms before the onset of the target noun for predictive sentences. For neutral sentences, the uninformative cue occurred 1012 ms prior to target noun onset, on average. Neutral sentences did not have a significantly greater time duration between the onset of uninformative verbs (e.g., see) and target nouns, as compared to the duration between the onset of the informative cues (e.g., little) and target nouns in predictive sentences (paired-sample $t(31) = 1.24$, $p = 0.226$).

Visual stimuli were photographic scene images of a table. Each scene image included singular and plural forms of four target objects: *apple, book, flower,* and *napkin* (Figure 5). These four objects appeared in different locations and with varying plurality across

Figure 5. Example of visual stimuli in Experiment 3. Participants viewed photographic stimuli and heard neutral sentences (e.g., *Do you see the red apples?*; *Can you find the red apples?*; *Can you show me the red apples?*) and various kinds of predictive sentences that included informative verbs, size adjectives, or number markings (e.g., *Could Sally eat the red apples?*; *Do you see the little red apples?*; *Where are the red apples?*; and *Can you see those red apples?*).

trials. Target objects had consistent colours, such that apples and flowers were always red and books and napkins were always blue. Each scene image was 1280 × 1024 pixels and appeared twice during the task (once with a neutral sentence and once with a predictive sentence).

Visual stimuli for Experiment 3 were therefore similar to Experiment 1 stimuli (i.e., photographic images with multiple referents) but were somewhat simpler. This simplification allowed listeners to generate accurate predictions based on multiple cues present in predictive sentences. For example, listeners could use number marking (e.g., *Where are the red apples?*) to rapidly and accurately predict a plural referent and to generate anticipatory eye movements to the appropriate object within the scene. Likewise, the colour adjectives also facilitated a narrowing of the scope of reference, but, as in Experiment 2, the colour adjectives did not uniquely identify the target referent.

During each trial, visual stimuli appeared for 2 s prior to the onset of auditory stimuli and remained visible for 3 s, such that the total duration of each trial was 5 s. The Experiment 3 preview time (2 s) was based on the intermediate complexity of the visual stimuli (four referents), as compared to Experiment 1 (many referents) and Experiment 2 (two referents), and was equal to prior developmental work which evaluated children's prediction abilities with four referents (Borovsky et al., 2012). Trials appeared in one of four quasi-randomized orders, which counterbalanced target plurality (singular or plural) and target location (upper right, upper left, lower right, or lower left), and ensured that condition (neutral or predictive), target plurality (singular or plural), and target object (apple, book, flower, or napkin) did not repeat for more than four trials sequentially. Filler trials occurred every eight trials, and consisted of a cartoon image (e.g., a smiling girl) and encouraging statements (e.g., "Amazing! You're almost done!"). In total, Experiment 3 included 16 predictive trials, 16 neutral trials, and 4 filler trials. All visual and auditory stimuli and experiment code for Experiment 3 are available on the Open Science Framework.

Results

As in Experiments 1 and 2, we analysed samples recorded within a 500 × 500 pixel area surrounding each visual referent and eliminated any samples that were outside of these visual areas of interest (759,554 of 5,494,093 samples, 14%) prior to aggregating data within 100-ms time-bins. We again analysed participants' looking behaviour during a time window from

1000 ms before to 1000 ms after the onset of the target noun. Our hypothesised results were identical to those of the prior experiments: If listeners use varied informative verbs, size adjectives, and number markings to predict the upcoming referent, then we expected them to generate anticipatory eye movements to the target referent, such that we could observe the emergence of condition effects *before* the onset of the target noun (0 ms).

Following the same procedures as in Experiments 1 and 2, we first analysed listeners' proportion of target looks during neutral and predictive sentences with a mixed-effects logistic regression model. Model results revealed significant effects for condition ($\beta = -0.09$, $z = -4.59$, $p < 0.001$), age group ($\beta = 0.38$, $z = 5.16$, $p < 0.001$), and time ($\beta = 2.43$, $z = 76.45$, $p < 0.001$), indicating that listeners' target looks were greater for predictive trials, that adults' target looks were greater than children's, and that listeners' target looks increased over time. As observed in Experiments 1 and 2, model results also revealed an interaction of age group and time ($\beta = 0.36$, $z = 11.30$, $p < 0.001$), indicating that adults' target looks increased at a faster rate than children's. Unlike the prior experiments, model results did not indicate a significant interaction of condition and time ($\beta = 0.06$, $z = 1.79$, $p = 0.074$). This marginally significant interaction suggests that, although listeners generated more target looks for predictive trials overall, changes in listeners' looking behaviour, over time, were similar across conditions in Experiment 3. Similarly, the three-way interaction of condition, age group, and time was not statistically significant, although the marginally significant result ($\beta = 0.06$, $z = 1.84$, $p = 0.066$) suggests that the marginally significant interaction of condition and time may have been more robust for adults than for children.

Next, as in the preceding experiments, we analysed adults' and children's looking behaviours with cluster-based permutation analyses. Findings revealed brief significant clusters which emerged prior to the onset of the target noun for adults (−300 to −100 ms, cluster $t = 4.92$, $p = 0.027$) and for children (−200 to 0 ms, cluster $t = 5.83$, $p = 0.005$). The observed condition effects in Experiment 3 were more attenuated relative to those observed in Experiments 1 and 2 (Figure 6).

As in Experiment 2, we next evaluated whether listeners used each type of predictive cue independently to anticipate upcoming information. We compared participants' proportion of target looks during each of the four types of predictive sentences (e.g., verbs: *Could Sally eat the red apples?*; size adjectives: *Do you see the little red apples?*; number markings: *Where are the red apples?*; and deictic number markings: *Can you see those red apples?*) with their proportion of target looks during neutral sentences. We analysed adults' and children's looking behaviours with mixed-effects logistic regression models, using the same specifications as in Experiment 2. Results for Experiment 3, summarised in Table 1, indicate that adults and children did not use all available types of informative linguistic cues to anticipate upcoming target nouns. There were no significant effects for number markings or deictic number markings for either age group, although the effect for adults was marginally significant. For size adjectives, effects were significant for adults but marginally significant for children. In contrast to these null and marginally significant effects, results showed robust prediction effects for informative verbs for both adults and children. Adding to these results, cluster-based permutation analyses only indicated significant clusters prior to the onset of the target noun for informative verbs (adults: −600 to 400 ms, cluster $t = 38.62$, $p < 0.001$; children: −500 to 600 ms, cluster $t = 46.98$, $p < 0.001$), suggesting that adults and children relied on verb semantics to generate predictions in Experiment 3. Results and figures for each predictive cue are available in Supplementary Materials on the Open Science Framework. In sum, Experiment 3 results converged with those of Experiment 2 by indicating robust significant effects for informative verbs, but results differed from Experiment 2 because there were null or marginal effects for size adjectives, number marking, and deictic number marking.

Comparisons of Experiments 2 and 3

While it was not possible to compare effects between Experiments 1 and 2 due to the divergent study designs, it was possible to directly compare results across Experiments 2 and 3. To do so, we analysed listeners' proportion of target looks during neutral and predictive sentences with a mixed-effects logistic regression model, including interacting fixed effects for experiment (Experiment 2, Experiment 3), condition (neutral, predictive), and time (100-ms bins, −1000 to 1000 ms from noun onset). The model also included random intercepts for subjects. Results revealed significant effects for condition ($\beta = 0.51$, $t = 12.58$, $p < 0.001$) and time ($\beta = 1.79$, $t = 37.87$, $p < 0.001$), as well as an interaction of condition and time ($\beta = -0.29$, $t = -4.34$, $p < 0.001$), respectively indicating that: listeners' target looks were greater overall for predictive trials, listeners' target looks increased over time, and listeners' target looks increased at a greater rate for predictive trials. Results also revealed a significant effect for experiment ($\beta = -1.05$, $t = -25.29$, $p < 0.001$), indicating that listeners generated more target looks in Experiment 2 than in Experiment 3. This

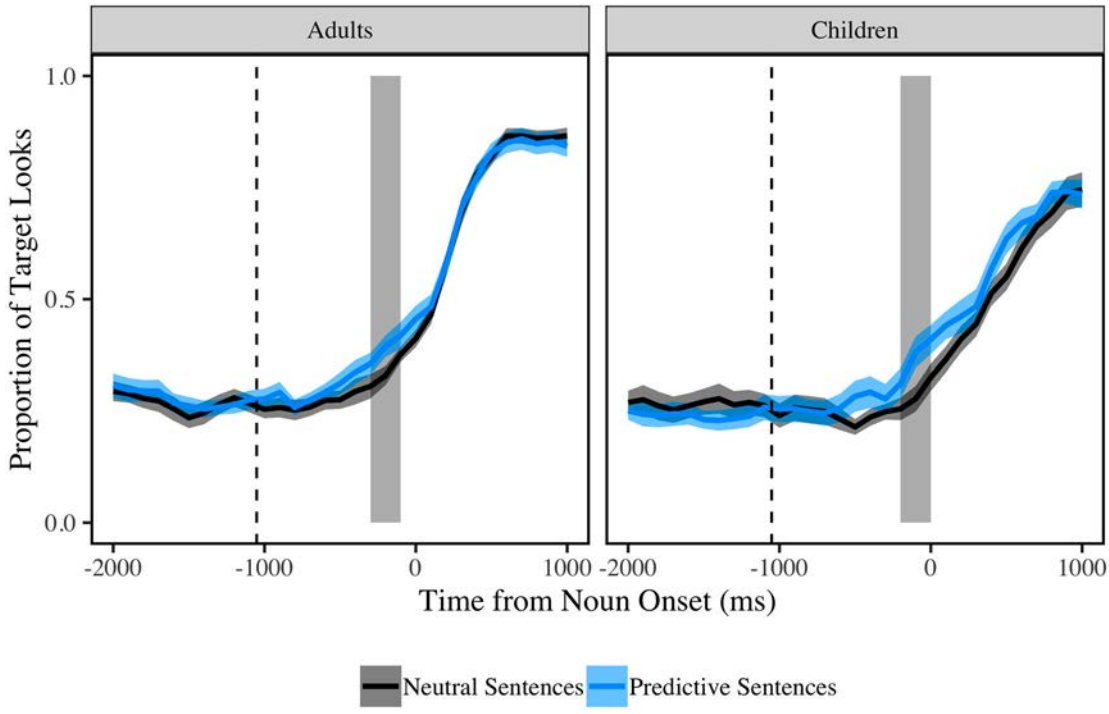

Figure 6. Looking-time plots for Experiment 3. Proportion of looks to the target referent during neutral sentences (grey) and predictive sentences (blue) for adults ($n = 24$) and for children ($n = 24$). The onset of the target noun (e.g., *apples*) is at 0 ms. Vertical dashed lines indicate the average onset of the informative word in predictive sentences (e.g., *eat*). Neutral and predictive sentences included a colour adjective (blue/red) immediately prior to the target noun, which partially narrowed the scope of reference prior to noun onset. Line shading indicates one standard error from the mean for each condition, averaged by subjects. Area shading indicates significant clusters ($ps < 0.05$) from permutation analyses (Maris & Oostenveld, 2007). Results indicate that adults and children generated anticipatory eye movements to the target referent during predictive sentences, but the condition effect in Experiment 3 was more attenuated relative to Experiments 1 and 2. Additional analyses suggest that listeners were primarily successful in using informative verbs to generate predictions in Experiment 3.

difference is to be expected, however, because Experiment 2 included only two visual referents, whereas Experiment 3 included four visual referents. Critically, results indicated a significant interaction of condition and experiment ($\beta = -0.38$, $t = -6.57$, $p < 0.001$), suggesting that condition differences (predictive > neutral) were more robust for Experiment 2 than for Experiment 3. Furthermore, results revealed a three-way interaction of condition, time, and experiment ($\beta = 0.27$, $t = 2.82$, $p < 0.001$), indicating that the Experiment 2 interaction effect for condition and time was more robust than Experiment 3. Together, these comparisons further substantiate the prior regression analyses and cluster analyses by indicating weaker condition effects in Experiment 3.

General discussion

To evaluate prediction in real-time language processing, prior developmental investigations have relied heavily on constrained experimental contexts with two-alternative visual referents and disproportionate exposure to a single linguistic cue across time. Therefore, it is unclear whether listeners are capable of generating predictions in response to the more complex and variable perceptual input that characterises real-world language processing circumstances. If prediction is a mechanism that supports children's language learning, then children must be able to generate predictions in more naturalistic communicative contexts. To make progress in addressing this limitation in prior research, we manipulated the complexity of the visual stimuli (Experiment 1), the variability of the speech stimuli (Experiment 2), and the complexity and variability of visual and speech stimuli simultaneously (Experiment 3). Findings indicate that both adults and children incorporated these somewhat naturalistic visual and auditory stimuli to rapidly and accurately generate predictions during real-time language processing. However, when both visual and auditory stimuli were complex and varied (Experiment 3), listeners did so less reliably, such that they only showed evidence of using informative verbs to generate predictions. This pattern of results suggests that listeners may be able to generate predictions within at least

mildly naturalistic communicative contexts, and if certain linguistic cues are available in naturalistic dialogue (i.e., informative verbs) then prediction may be a viable developmental mechanism.

The present findings converge with and extend prior findings in a number of ways. Experiment 1 indicated that adults and children can generate predictions in response to complex visual stimuli (i.e., photographic images). These results provide a conceptual replication of prior findings with adult listeners (Coco et al., 2016; Staub et al., 2012) and extend those findings by indicating that children, in a manner similar to adults, can engage in predictive processing while navigating somewhat naturalistic visual scenes. Experiment 2 converged with a number of prior developmental investigations by indicating that adults and children can use informative verbs (Fernald et al., 2008; Mani & Huettig, 2012), adjectives (Fernald et al., 2010), and number markings (Lew-Williams, 2017; Lukyanenko & Fisher, 2016; Reuter et al., 2020) as a basis for prediction. These results suggest that listeners can keep pace with variable speech input over time, at least when the visual scene is not cluttered. Experiment 3 further suggests that both adults and children may be able to generate predictions when visual and speech information is complex and variable, but only for a subset of semantic cues.

The present findings lend support to theories that prioritize prediction as a mechanism for children's language learning (Chater et al., 2016; Dell & Chang, 2014; Elman, 1990, 2009), but they do not provide conclusive evidence that prediction occurs in real-world language processing contexts. Therefore, the findings must be interpreted with caution. In particular, by comparing the results of Experiments 2 and 3, we find that prediction effects which are robust in a simple visual context, such as number markings (Lew-Williams, 2017; Lukyanenko & Fisher, 2016; Reuter et al., 2020) are absent in a more complex visual context. But even these non-significant effects must also be interpreted with caution. It is possible that listeners use number markings in real-world communicative contexts to generate predictions but not in constrained lab contexts. Further research is needed to characterize what predictive cues are regularly available in learners' day-to-day environments and how the dynamics of prediction may vary between lab and real-life processing contexts. That said, our experiments do provide possible nuance about how prediction may occur in nature. Specifically, listeners may exploit some sources of information (i.e., verb semantics) but not other sources of information (i.e., adjectives and number markings) for generating predictions. In controlled settings but also likely in natural settings, verb semantics may be the primary basis for listeners' real-time predictions. Thus, beyond determining *whether* prediction might occur, the present findings highlight the need to investigate *how* different forms of prediction may occur in naturalistic dialogue – both their diversity and reliability.

Numerous limitations will need to be addressed in order to further specify whether and how prediction supports language processing and language development in everyday conversation. Notably, the present experiments tested 4- and 5-year-old children, and it is possible that younger listeners may lack the necessary language experience or cognitive resources to rapidly and accurately generate predictions in varied language processing contexts (Pickering & Gambi, 2018; Rabagliati et al., 2016). Thus, future developmental investigations must incorporate participants across a broader age-range, with a focus on 1- to 3-year-old children who are just breaking into the sound sequences, words, and sentences of their ambient language(s). A second limitation of the present investigation is that the experimental contexts retained a moderate level of visual and auditory constraint. The full ecological validity of the present findings therefore remains uncertain. However, the present design represents an important step toward evaluating prediction within increasingly natural processing contexts. Next steps in research could use head-mounted eye-tracking methods to assess whether and how prediction occurs during unscripted, day-to-day conversations (Tanenhaus & Brown-Schmidt, 2008). Indeed, prior findings suggest that, although their visual input is complex and cluttered, infants could plausibly use a small set of consistent referents, such as spoons at mealtimes, as a basis for predicting and learning (Clerkin et al., 2017). Future studies could also make use of naturalistic language corpora (e.g., VanDam et al., 2016) and video-based corpora to evaluate the extent to which different linguistic cues such as verbs, adjectives, and number markings could support prediction in real-world language processing contexts. Relatedly, further work could determine the extent to which predictions are generated across diverse sentence constructions. For instance, verb class and argument structure may play a role in shaping listeners' predictions both in lab-based tasks and in more naturalistic contexts. Do listeners accurately predict upcoming referents in sentences with dative alternations (e.g., "The girl gave the boy a letter" vs. "The girl gave the letter to the boy")? Does prediction occur only for a limited set of verbs and constructions, and if so, are these available in learners' everyday linguistic experiences? Finally, while the present study relied primarily on semantic and syntactic cues for predictions,

there are many other sources of information that children could use for generating predictions, including paralinguistic cues such as speech disfluencies and speaker identity (Borovsky & Creel, 2014; Bosker et al., 2014; Creel, 2012, 2014; Kidd et al., 2011). Finally, the preview times for visual stimuli varied across our experiments, reflecting different traditions in language processing research, and preview times may have differentially influenced listeners' predictions (Huettig & Guerra, 2019). To further flesh out the nature of real-time prediction in natural contexts, we will need to systematically determine how recent visual experience interacts with listeners' propensities to initiate predictions.

In sum, the present investigation took a step toward addressing the ecological validity of previous findings on children's abilities to generate predictions during real-time language processing. Findings broadly suggest that both adults and children are capable of generating predictions in visually complex scenes or when the available linguistic cues vary from one sentence to the next. When visual and speech stimuli are both complex and variable, the scope of adults' and children's predictions may be somewhat narrower, such that they retain the ability to exploit some linguistic cues (e.g., informative verbs) more than others (e.g., informative adjectives or number markings). Overall, this pattern of results lends modest support to the idea that prediction is a viable developmental mechanism that supports processing and learning. Perceptual variability is an important aspect of natural communicative contexts, but further work is needed to explore whether, when, and how prediction supports development in the full dimensionality of children's everyday interactions with others.

Notes

1. However, it is important to distinguish behavioral measures of prediction from prediction itself: Although prediction is typically operationalized via anticipatory eye movements, prediction may occur before or in the absence of overt behaviours. For example, a listener might accurately predict a speaker's referent before they visually locate it within the surrounding scene. Similarly, listeners could presumably anticipate an abstract or absent referent.

Acknowledgements

We thank all participants, as well as Dominick Reuter, Claire Robertson, and Cynthia Lukyanenko for assistance with stimuli, Mia Sullivan for assistance with data collection, and other members of the Princeton Baby Lab for assistance with participant recruitment. We are also grateful to Adele Goldberg for comments on a previous version of this paper. This research was supported by grants from the National Institute of Child Health and Human Development to Casey Lew-Williams (R01HD095912, R03HD079779) and from the National Science Foundation to Tracy Reuter (DGE-1656466).

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by grants from the National Institute of Child Health and Human Development to CLW [R01HD095912,R03HD079779] and from the National Science Foundation to TR [DGE1656466].

ORCID

Tracy Reuter http://orcid.org/0000-0003-2714-0153

References

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*(3), 247–264. https://doi.org/10.1016/s0010-0277(99)00059-1

Altmann, G. T. M., & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language*, *57*(4), 502–518. https://doi.org/10.1016/j.jml.2006.12.004

Andersson, R., Ferreira, F., & Henderson, J. M. (2011). I see what you're saying: The integration of complex speech and scenes during language comprehension. *Acta Psychologica*, *137*(2), 208–216. https://doi.org/10.1016/j.actpsy.2011.01.007

Andreu, L., Sanz-Torrent, M., & Trueswell, J. C. (2013). Anticipatory sentence processing in children with specific language impairment: Evidence from eye movements during listening. *Applied Psycholinguistics*, *34*(1), 5–44. https://doi.org/10.1017/S0142716411000592

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Barthel, M., Meyer, A. S., & Levinson, S. C. (2017). Next speakers plan their turn early and speak after turn-final "go-signals". *Frontiers in Psychology*, *8*, 1–10. https://doi.org/10.3389/fpsyg.2017.00393

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

Bobb, S. C., Huettig, F., & Mani, N. (2016). Predicting visual information during sentence processing: Toddlers activate an object's shape before it is mentioned. *Journal of Experimental Child Psychology*, *151*, 51–64. https://doi.org/10.1016/j.jecp.2015.11.002

Boersma, P., & Weenink, D. (2017). *Praat: doing phonetics by computer* (Version 6.0.19). 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. https://doi.org/10.1016/j.cognition.2020.104347

Borovsky, A., Burns, E., Elman, J. L., & Evans, J. L. (2013). Lexical activation during sentence comprehension in adolescents with history of specific language impairment. *Journal of Communication Disorders*, *46*(5–6), 413–427. https://doi.org/10.1016/j.jcomdis.2013.09.001

Borovsky, A., & Creel, S. (2014). Children and adults integrate talker and verb information in online processing. *Developmental Psychology*, *50*(5), 1600–1613. https://doi.org/10.1037/a0035591

Borovsky, A., Elman, J. L., & 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://doi.org/10.1016/j.jecp.2012.01.005

Bosker, H. R., Quené, H., Sanders, T., & de Jong, N. H. (2014). Native 'um's elicit prediction of low-frequency referents, but non-native 'um's do not. *Journal of Memory and Language*, *75*, 104–116. https://doi.org/10.1016/j.jml.2014.05.004

Brouwer, S., Mitterer, H., & Huettig, F. (2013). Discourse context and the recognition of reduced and canonical spoken words. *Applied Psycholinguistics*, *34*(3), 519–539. https://doi.org/10.1017/S0142716411000853

Chater, N., McCauley, S. M., & Christiansen, M. H. (2016). Language as skill: Intertwining comprehension and production. *Journal of Memory and Language*, *89*, 244–254. https://doi.org/10.1016/j.jml.2015.11.004

Clerkin, E. M., Hart, E., Rehg, J. M., Yu, C., & Smith, L. B. (2017). Real-world visual statistics and infants' first-learned object names. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1711), 20160055. https://doi.org/10.1098/rstb.2016.0055

Coco, M. I., Keller, F., & Malcolm, G. L. (2016). Anticipation in real-world scenes: The role of visual context and visual memory. *Cognitive Science*, *40*(8), 1995–2024. https://doi.org/10.1111/cogs.12313

Creel, S. C. (2012). Preschoolers' use of talker information in online comprehension. *Child Development*, *83*(6), 2042–2056. https://doi.org/10.1111/j.1467-8624.2012.01816.x

Creel, S. C. (2014). Preschoolers' flexible use of talker information during word learning. *Journal of Memory and Language*, *73*(1), 81–98. https://doi.org/10.1016/j.jml.2014.03.001

Dell, G. S., & Chang, F. (2014). The P-chain: Relating sentence production and its disorders to comprehension and acquisition. *Philosophical Transactions of the Royal Society B*, *369*, 1–8. https://doi.org/10.1098/rstb.2012.0394

Eberhard, K. M., Spivey-Knowlton, M. J., Sedivy, J. C., & Tanenhaus, M. K. (1995). Eye movements as a window into real-time spoken language comprehension in natural contexts. *Journal of Psycholinguistic Research*, *24*(6), 409–436. https://doi.org/10.1007/BF02143160

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, *14*(2), 179–211. https://doi.org/10.1016/0364-0213(90)90002-E

Elman, J. L. (2009). On the meaning of words and dinosaur bones: Lexical knowledge without a lexicon. *Cognitive Science*, *33*(4), 547–582. https://doi.org/10.1111%2Fj.1551-6709.2009.01023.x

Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology*, *44*(4), 491–505. https://doi.org/10.1111/j.1469-8986.2007.00531.x

Fernald, A., Thorpe, K., & Marchman, V. A. (2010). Blue car, red car: Developing efficiency in online interpretation of adjective-noun phrases. *Cognitive Psychology*, *60*(3), 190–217. https://doi.org/10.1075/lald.44.06fer

Fernald, A., Zangl, R., Portillo, A. L., & Marchman, V. A. (2008). Looking while listening: Using eye movements to monitor spoken language comprehension by infants and young children. In I. A. Sekerina, E. M. Fernández, & H. Clahsen (Eds.), *Developmental psycholinguistics: On-line methods in children's language processing* (pp. 97–135). John Benjamins. https://doi.org/10.1075/lald.44

Gambi, C., Gorrie, F., Pickering, M. J., & Rabagliati, H. (2018). The development of linguistic prediction: Predictions of sound and meaning in 2-to-5 year olds. *Journal of Experimental Child Psychology*, *44*(0), 1–54. https://doi.org/10.1016/j.jecp.2018.04.012

Gibson, E., Bergen, L., & Piantadosi, S. T. (2013). Rational integration of noisy evidence and prior semantic expectations in sentence interpretation. *Proceedings of the National Academy of Sciences*, *110*(20), 8051–8056. https://doi.org/10.1073/pnas.1216438110

Havron, N., de Carvalho, A., Fiévet, A. C., & Christophe, A. (2019). Three-to four-year-old children rapidly adapt their predictions and use them to learn novel word meanings. *Child Development*, *90*(1), 82–90. https://doi.org/10.1111/cdev.13113

Huettig, F. (2015). Four central questions about prediction in language processing. *Brain Research*, *1626*, 118–135. https://doi.org/10.1016/j.brainres.2015.02.014

Huettig, F., & Guerra, E. (2019). Effects of speech rate, preview time of visual context, and participant instructions reveal strong limits on prediction in language processing. *Brain Research*, *1706*, 196–208. https://doi.org/10.1016/j.brainres.2018.11.013

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

Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, *137*(2), 151–171. https://doi.org/10.1016/j.actpsy.2010.11.003

Ito, A., Corley, M., & Pickering, M. J. (2018). A cognitive load delays predictive eye movements similarly during L1 and L2 comprehension. *Bilingualism: Language and Cognition*, *21*(2), 251–264. https://doi.org/10.1017/S1366728917000050

Kamide, Y. (2008). Anticipatory processes in sentence processing. *Language and Linguistics Compass*, *2*(4), 647–670. https://doi.org/10.1111/j.1749-818X.2008.00072.x

Kidd, C., White, K. S., & Aslin, R. N. (2011). Toddlers use speech disfluencies to predict speakers' referential intentions. *Developmental Science*, *14*(4), 925–934. https://doi.org/10.1111/j.1467-7687.2011.01049.x

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

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. https://doi.org/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://doi.org/10.1111/j.1467-9280.2007.01871.x

Lew-Williams, C., & Fernald, A. (2009). Fluency in using morpho-syntactic cues to establish reference: How do native and non-native speakers differ? In J. Chandlee, M. Franchini, S. Lord, & G. Rheiner (Eds.), *Proceedings of the 33rd annual Boston university conference on language development* (pp. 290–301). Cascadilla Press.

Luck, S. J. (2012). Electrophysiological correlates of the focusing of attention within complex visual scenes: N2pc and related ERP components. In S. J. Luck & E. S. Kappenman (Eds.), *Oxford library of psychology: The Oxford handbook of event-related potential components* (pp. 329–360). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195374148.013.0161

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. https://doi.org/10.1016/j.cognition.2015.10.012

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*(4), 843–847. https://doi.org/10.1037/a0029284

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

Nation, K., Marshall, C. M., & Altmann, G. T. (2003). Investigating individual differences in children's real-time sentence comprehension using language-mediated eye movements. *Journal of Experimental Child Psychology*, *86*(4), 314–329. https://doi.org/10.1016/j.jecp.2003.09.001

Nencheva, M. L., Piazza, E. A., & Lew-Williams, C. (2020). The moment-to-moment pitch dynamics of child-directed speech shape toddlers'attention and learning. *Developmental Science*. https://doi.org/10.1111/desc.12997

Otten, M., & Van Berkum, J. J. A. (2009). Does working memory capacity affect the ability to predict upcoming words in discourse? *Brain Research*, *1291*, 92–101. https://doi.org/10.1016/j.brainres.2009.07.042

Pickering, M. J., & Gambi, C. (2018). Predicting while comprehending language: A theory and review. *Psychological Bulletin*, *144*(10), 1002–1044. https://doi.org/10.1037/bul0000158

Pickering, M. J., & Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences*, *11*(3), 105–110. https://doi.org/10.1016/j.tics.2006.12.002

Rabagliati, H., Gambi, C., & Pickering, M. J. (2016). Learning to predict or predicting to learn? *Language, Cognition and Neuroscience*, *31*(1), 94–105. https://doi.org/10.1080/23273798.2015.1077979

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

Reuter, T., Sullivan, M., & Lew-Williams, C. (2020). *Look at that: Spatial deixis reveals experience-related differences in prediction*. [Manuscript submitted for publication]

Snedeker, J., & Trueswell, J. C. (2004). The developing constraints on parsing decisions: The role of lexical-biases and referential scenes in child and adult sentence processing. *Cognitive Psychology*, *49*(3), 238–299. https://doi.org/10.1016/j.cogpsych.2004.03.001

Sorensen, D. W., & Bailey, K. G. D. (2007). The world is too much: Effects of array size on the link between language comprehension and eye movements. *Visual Cognition*, *15*, 112–115. https://doi.org/10.1080/13506280600975486

Staub, A., Abbott, M., & Bogartz, R. S. (2012). Linguistically guided anticipatory eye movements in scene viewing. *Visual Cognition*, *20*(8), 922–946. https://doi.org/10.1080/13506285.2012.715599

Tanenhaus, M. K., & Brown-Schmidt, S. (2008). Language processing in the natural world. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1493), 1105–1122. https://doi.org/10.1098/rstb.2007.2162

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*(5217), 1632–1634. https://doi.org/10.1126/science.7777863

VanDam, M., Warlaumont, A. S., Bergelson, E., Cristia, A., Soderstrom, M., De Palma, P., & MacWhinney, B. (2016). Homebank: An online repository of daylong child-centered audio recordings. *Seminars in Speech and Language*, *37*(2), 128–142. https://doi.org/10.1055/s-0036-1580745

Wittenberg, E., Khan, M., & Snedeker, J. (2017). Investigating thematic roles through implicit learning: Evidence from light verb constructions. *Frontiers in Psychology*, *8*, 1–8. https://doi.org/10.3389/fpsyg.2017.0108

Yurovsky, D., Case, S., & Frank, M. C. (2017). Preschoolers flexibly adapt to linguistic input in a noisy channel. *Psychological Science*, *28*(1), 132–140. https://doi.org/10.1177/0956797616668557