The relation between preschoolers' vocabulary growth and their ability to predict and recognize words.

Chiara Gambi

University of Edinburgh and Cardiff University

Priya Jindal

Sophie Sharpe

Martin J. Pickering

Hugh Rabagliati

University of Edinburgh

--Final author manuscript *in press* in Child Development, 2020--

Address for correspondence:

Chiara Gambi

School of Psychology

70, Park Place

Cardiff University

CF10 3AT Cardiff, U.K.

GambiC@cardiff.ac.uk

Phone: +44(0)29 206 88950

Abstract

By age 2, children are developing foundational language processing skills, such as quickly recognizing words and predicting words before they occur. How do these skills relate to children's structural knowledge of vocabulary? Multiple aspects of language processing were simultaneously measured in a sample of 2-to-5-year-olds (N=215): While older children were more fluent at recognizing words, at predicting words in a graded fashion, and at revising incorrect predictions, only revision was associated with concurrent vocabulary knowledge once age was accounted for. However, an exploratory longitudinal follow-up (N=55) then found that word recognition and prediction skills were associated with rate of subsequent vocabulary development, but revision skills were not. We argue that prediction skills may facilitate language learning through enhancing processing speed.

Keywords: vocabulary development; linguistic prediction; word recognition; eye-tracking; longitudinal

The relation between preschoolers' vocabulary development and their ability to predict and recognize words.

Children show considerable variation in how quickly they acquire knowledge about their native language(s), e.g., about the structure and composition of their vocabulary (Fenson et al., 1994). While there is strong evidence that this variation can be partially predicted by environmental factors, such as quantity and quality of early linguistic input (e.g., Hiareau, Yeung, & Nazzi, 2019; Hoff, 2003; Huttenlocher, Haight, Bryk, Seltzer, & Lyons, 1991; Rowe, 2012; Weisleder & Fernald, 2013; Weizman & Snow, 2001), recent work also suggests how certain child-internal factors may play an important explanatory role. Of particular interest here, children's ability to efficiently process linguistic input, such as quickly recognizing words and grasping sentence meaning, has been robustly associated with their concurrent vocabulary knowledge, and also with later language outcomes (Fernald, Perfors, & Marchman, 2006; Fernald & Marchman, 2012; Marchman & Fernald, 2008; Peter, et al., 2019; Weisleder & Fernald, 2013; see also Duff, Reen, Plunkett, & Nation, 2015; Friend, Smolak, Liu, Poulin-Dubois, & Zesiger, 2018 for evidence that current vocabulary also predicts later language outcomes). But what is the relation between children's ability to *predict* upcoming linguistic input and their concurrent and later vocabulary knowledge?

Links between language processing skills and language outcomes are expected under a variety of theories of language development, all incorporating the idea that the way in which children process and make sense of their linguistic input in-the-moment shapes what and how much they can learn from it (McCauley & Christiansen, 2019; Omaki & Lidz, 2015; Pozzan & Trueswell, 2015). Here, we focus in particular on the kind of relation that is expected under models of error-driven learning (Chang, Dell, & Bock, 2006; Ramscar, Dye, & McCauley, 2013). In such models, children learn about meaning and grammar by continuously predicting what they will hear next based on their current knowledge of how words are used, and revising that knowledge when their predictions are incorrect.

As we describe below, there is considerable evidence that children predict upcoming words when processing sentences (Borovsky, Elman, & Fernald, 2012; Gambi, Pickering, & Rabagliati, 2016; Mani & Huettig, 2012), and these models therefore assume that there should be a particularly strong relation between children's language outcomes and their skill at predicting linguistic input. In this context, prediction skill is a measure of children's ability to generate expectations about the words they will encounter, before they encounter them, and it contrasts with recognition skill, a measure of how quickly children can access the meaning of a spoken word as they hear it (Pickering & Gambi, 2018). Here, we assess whether pre-schoolers' prediction skills relate to both their concurrent vocabulary size and longitudinal vocabulary development; furthermore, in the same children, we assess the relations between recognition skills and concurrent and later vocabulary knowledge (Fernald, et al., 2006). The aim is to investigate both *whether* and *how* prediction skill may be related to the development of linguistic knowledge.

How might prediction relate to language learning?

By their second birthday, children begin to develop an increasingly sophisticated ability to predict upcoming language. For example, two-year-olds can already use the meaning of a known verb to predict a likely object (e.g., *eat* predicts *apple*; Mani, Daum, & Huettig, 2016; Mani & Huettig, 2012). From the age of 3, children begin to combine semantic associations elicited by the subject and verb of a transitive sentence to predict the most appropriate continuation (e.g., *pirate* plus *chase* predicts *ship*, but *dog* plus *chase* predicts *cat*; Borovsky et al., 2012). Moreover, preschoolers are also able to combine meaning and grammar, so that they predict strong semantic associates only if they fulfill an available grammatical role (e.g., *Mary will arrest the...* predicts *robber*, but not *policeman*, because the agent role is not available; Gambi et al., 2016). In sum, when children generate predictions about upcoming words, they make use of all of their developing linguistic knowledge, and are clearly able to anticipate the most likely continuation of transitive verb frames.

These skills at prediction could be related to language development because prediction facilitates language learning, and this facilitation could come about in one of two ways (Rabagliati, Gambi, & Pickering, 2015). Under error-driven learning models of language development, prediction plays a key role in the process of learning: Children are assumed to continuously generate predictions about upcoming language, and they learn by comparing these predictions to the input, which generates informative error signals, and triggers updating of their internal language model (Chang et al., 2006; Ramscar et al., 2013). Thus, under these models, children's prediction skills play a direct role in their linguistic development. In contrast, under other models of language learning, prediction may still play an important role, but it would do so indirectly, through the facilitative effect that prediction exerts on fluent language processing (Fernald, Marchman, & Hurtado, 2008; Omaki & Lidz, 2015; Pozzan & Trueswell, 2015). As Fernald and colleagues argue (Fernald, Marchman, et al., 2008), children who can quickly and fluently process the linguistic and non-linguistic context around a novel word are at an advantage in trying to guess what the speaker intends it to mean. Prediction can enhance fluent processing because it permits predictable words to be pre-processed, and thus speeds up recognition times (Lew-Williams & Fernald, 2007; Mahr, McMillan, Saffran, Weismer, & Edwards, 2015). Attentional resources can therefore be devoted elsewhere, such as to more accurately infer the meanings of novel words using linguistic and non-linguistic cues.

Consistent with both of these ideas, recent evidence does suggest a relation between children's skill at prediction and their language-learning outcomes. For example, 3-to-4-year-olds' predictions about how people use ambiguous syntactic frames affect what word meanings they learn. When primed to interpret an ambiguous frame (e.g., French *la petite*) as a noun (i.e., "the small one" vs. an adjective: "the small"), children learned action meanings for novel words inserted after the frame (*la petite dase*), presumably because they predicted that a verb would follow the noun (Havron, de Carvalho, Fiévet, & Christophe, 2019). Further, 3-to-5 year olds' ability to reorient after an incorrect prediction correlates with their skill at learning novel words (Reuter, Borovsky, & Lew-Wlliams, 2019). In an eye-tracking task, children heard sentences like *Yummy, let's eat soup! I'll stir it with a cheem*, where the context predicts *spoon* but *cheem* referred to a novel tool. Reuter and colleagues found that children who showed evidence of learning the novel words were more likely to engage in a predict-and-redirect strategy, initially predicting (gazing towards) a depicted spoon while listening to the context, but then quickly re-orienting their gaze towards the novel tool when they heard *cheem*. Finally, there is evidence that children's skill at predicting words while listening to sentences correlates with their current linguistic knowledge, particularly their vocabulary size, both for preschool and school-age children (Borovsky et al., 2012), and for children as young as 24 months (Mani & Huettig, 2012).

However, while these findings are suggestive of a relation between prediction and learning, they are not conclusive about the nature and strength of that relation. First, much of the evidence is consistent with both accounts of how prediction facilitates learning: For example, the fact that structural predictions shape children's word learning (Havron et al., 2019) can be explained both by models in which prediction affects learning directly, via the computation of error signals, and by models in which it affects learning indirectly, because it facilitates fluent language processing and ambiguity resolution. Similarly, the finding that children's ability to reorient after an incorrect prediction is important for word learning (Reuter et al., 2019) could be explained in different ways: It could indicate a direct causal relation between error-revision and learning, or it could be that general cognitive ability means that children who are stronger learners are also better at revising incorrect predictions.

In addition, it is unclear to what extent young children would be able to learn from generating expectations that turn out to be incorrect. Specifically, this idea seems at odds with a large literature showing that, in many linguistic contexts, children struggle to revise their initial interpretations of sentences even at the end of the preschool years (Choi & Trueswell, 2010; Huang, Zheng, Meng, & Snedeker, 2013; Trueswell, Sekerina, Hill, & Logrip, 1999; Leech, Rowe, & Huang, 2017). If children's revision skills develop slowly, and thus they have difficulty updating their linguistic knowledge in real-time, then the influence of error-driven learning mechanisms in early development may be limited. Indeed, there is evidence that children who initially generate an incorrect hypothesis during a word learning task fail to encode information that could help them revise their incorrect hypothesis and arrive at the correct knowledge (Woodard, Gleitman, & Trueswell, 2016; Aravind, de Villiers, Pace, Valentine, Golinkoff, Hirsh-Pasek, … , & Wilson, 2018; but see Roembke & McMurray, 2016). Furthermore, revision difficulties also call into question the claim that prediction facilitates learning by enhancing fluent processing. In particular, processing delays due to incorrect predictions may well outweigh the speed up in recognition times that children experience when their predictions are correct (Omaki & Lidz, 2015), making the idea that prediction facilitates children's fluent language processing also a potentially problematic one.

Finally, while there is evidence of a relation between prediction skill and concurrent language knowledge, that evidence is surprisingly fragile. For example, while Mani and Huettig (2012) found that prediction skill did correlate with expressive vocabulary, it did not correlate with receptive vocabulary in the same sample, even though prediction skill did correlate with receptive vocabulary in older children (Borovsky et al., 2012). Further, in two studies, Gambi and colleagues found no evidence that prediction skill correlated with either productive or receptive vocabulary size in pre-schoolers, once age was controlled for (Gambi, Gorrie, Pickering, & Rabagliati, 2018; Gambi et al., 2016). Finally, the evidence that would be most informative – a longitudinal relation between prediction skill and later language outcomes – is yet to be collected. In the absence of such evidence, it is possible that these associations between prediction skills and linguistic knowledge arise because more linguistically advanced children are also better equipped to generate predictions - i.e., because prediction is a result of linguistic development, rather than because prediction plays a role in linguistic development (Rabagliati et al., 2015). In contrast, there is strong evidence for a relation between linguistic processing speed, as measured by how quickly children recognize spoken words (i.e., recognition skill), and both concurrent and later language outcomes (Fernald, Marchman, et al., 2008; Fernald, et al., 2006; Marchman & Fernald, 2008).

In sum, the evidence for a relation between prediction skills and vocabulary development is suggestive but not conclusive and, furthermore, we are yet to establish how and why prediction skill might be related to linguistic development: Does prediction facilitate language development in-and-of itself (e.g., via error-driven learning), or does it simply contribute to the broader facilitative effect of faster language processing? In order to address these questions, we not only need more robust evidence for a relation between prediction skill and both concurrent and later vocabulary knowledge, but also a better measurement of the degree of sophistication of young children’s ability to generate and revise linguistic expectations. Finally, we need to measure such prediction and revision skills alongside general word processing skills in order to understand how they jointly contribute to vocabulary development.

The current study

In the present work we aimed to understand whether and how children's linguistic prediction skills are associated with vocabulary knowledge and vocabulary development. To do this, we developed a visual world eye-tracking task that measured the sophistication of children's ability to predict upcoming words by assessing gradedness, that is the extent to which children can predict several alternative continuations, each in proportion to its degree of predictability; for example, predicting the most likely word very strongly, but also predicting a less likely word more strongly than a completely implausible word.

Capturing the gradedness of predictions is important both theoretically and methodologically. Graded predictions appear to be characteristic of adult language processing; for instance, on the basis of a timed sentence completion task, Staub and colleagues (Staub, Grant, Astheimer, & Cohen, 2015) showed that adults activate many possible continuations in parallel (see also Carter, Foster, Muncy, & Luke, 2019; Luke & Christianson, 2016; Smith & Levy, 2013) Thus, since expert language users predict in a highly graded fashion, we would expect children whose predictions are more graded (and thus more adult-like), to be more linguistically advanced. Accordingly, Mani et al. (2016) found that two-year-olds with larger expressive vocabularies were more likely to predict both words strongly associated with a sentence context and words that were only weakly associated with it, compared to an unassociated word. But while this suggests a relation between graded predictions and linguistic ability, the same study also found no relation between children's expressive vocabulary and the degree to which they predicted strong associates more than weak associates. Thus, more evidence is needed as to how the gradedness of children's predictions relates to their vocabulary knowledge.

In addition, we suggest that a measure of the gradedness of predictions is likely to have discriminative measurement properties that are useful for an individual differences design. One reason why evidence for a relation between prediction skills and linguistic knowledge has so far been inconsistent may be that measures of prediction skill have typically been limited to the child's ability to predict a single, highly predictable alternative (Borovsky et al., 2012; Gambi et al., 2016; Mani & Huettig, 2012). A more fine-grained assessment of gradedness, characterising the child's ability to distinguish between multiple differentially predictable alternatives, may provide a more sensitive measure of individual differences in linguistic prediction skill.

In our design, children heard sentences while viewing pictures that were differentially likely to be the final word (e.g., seeing a bone, slippers and pyjamas while hearing *Alfie's dog likes to chew on the…. bone*, where *bone* is more likely than *slippers*, and *slippers* is in turn more likely than *pyjamas* prior to hearing the final word). An advantage of this design is that it could naturally be extended to measure and test other factors. First, by including neutral, non-predictive sentences (e.g., *Now, Craig is looking for the bone*) we could measure the efficacy of children's language processing by capturing the speed with which they recognize spoken words without contextual facilitation (Fernald et al., 2006). Second, by varying the final word heard, we could measure children's responses to errors of prediction, capturing the degree to which they can quickly update their comprehension when their predictions are incorrect (Reuter et al., 2019). In particular, we compared word recognition times following neutral sentence contexts, when the final word was no more or less predictable than other options, to word recognition times when the final word was less predictable than a competitor, e.g., comparing recognition of *slippers* in *Now, Craig is looking for the slippers* (a neutral context), to *Alfie's dog likes to chew on the slippers,* where the competitor *bone* is more predictable than *slippers*. If children have difficulty revising following errors of prediction, then we would expect word recognition to proceed more slowly in the presence of a more predictable competitor.

We then assessed how these three measures – of prediction skill, processing speed, and revision skill – related to children's vocabulary development. Initially, we did this synchronously, and assessed how the three processing skills related to concurrent receptive vocabulary size in a large sample (N=215) of children aged 2-5 years (Phase 1). Then, seven months later (on average), we re-assessed the vocabulary size of a smaller opportunity sample of these children (N=55), which allowed us to conduct additional, exploratory analyses of how these same processing skills predicted subsequent change in vocabulary size (Phase 2).

Specifically, these exploratory analyses allowed us to assess whether our longitudinal data were more consistent with one of two competing hypotheses regarding the relation between prediction-related processing skills (including both prediction skill and revision skill) and vocabulary development. According to the first hypothesis, prediction facilitates language development in-and-of itself, and so we would expect to find that prediction-related processing skills explain variance in vocabulary development over and above measures of processing speed. In contrast, the second hypotheses maintains that prediction facilitates language development because it contributes to faster language processing, so we would expect prediction-related processing skills and measures of word processing speed to explain largely overlapping variance in vocabulary development.

Methods

For reasons of space and clarity, ancillary details of our methods, as well as additional analyses, can be found in the Supplementary Materials. Supplement sections are marked with a *§*. All data and analysis scripts are available at https://osf.io/9ckwe/.

Participants

Testing took place in two phases. For Phase 1 (April-June 2016), we did not conduct a formal power analysis, but rather based our data collection targets on previous eye-tracking studies of linguistic prediction in pre-schoolers (e.g., 40-47 children in each of 3 age groups in Gambi et al., 2018; 72 children in Gambi et al., 2016; 48 children in Borovsky et al., 2012; 30 children in Mani and Huettig, 2012 and in Mani et al., 2016). Our final sample size was larger than any of these previous studies (total N = 215): We tested 60 English-speaking two-year-olds (M_{age}: 30 months, range [24,35], 32 males), 77 three-year-olds (M_{age}: 41 months, range [36,47], 50 males), and 78 four-to-five-year-olds (M_{age}: 54 months, range [48,65], 32 males) in our lab (24 children) or at nursery schools in and around Edinburgh. Nine more children's data were discarded because of equipment malfunction (3), experimenter error (1), speech delay (2), or fussiness (3).

In Phase 2 (November 2016-February 2017), an opportunistic sub-sample of 55 children was retested (32 males; M_{age} at first test: 42 months, range [25, 60]; M_{age} at retest: 50 months, range [34, 68]) after a 5-to-10 months delay (M = 7.4 months, SD = 1.2). Phase 2 was not planned until after the end of Phase 1, hence the variability in the duration of the test-retest delay across children. One additional child's data was discarded because they had been excluded from Phase 1. We did not collect socio-economic status (SES) information for the full sample; however, we did collect it for the sub-sample. Our SES measure was the Scottish Index of Multiple Deprivation (*Scottish Index of Multiple Deprivation - SIMD16 Technical Notes*, 2016), with each child being assigned to the vigintile corresponding to their home postcode; for correlations between SES and processing and linguistic knowledge measures, see Supplementary Materials, *§3*. Children came predominantly from white, mid-to-high SES families.

INSERT FIGURE 1 HERE

INSERT TABLE 1 HERE

Materials and Procedure

In Phase 1, children completed a visual-world eye tracking task that assessed gradedness of predictions, revision skill, and processing speed. Then, they completed an assessment of receptive vocabulary (the British Picture Vocabulary Scale, BPVS; Second Edition, Dunn, Dunn, Whetton, & Burley, 1997). In Phase 2, children first completed the Test for Reception of Grammar (TROG; Second Edition, Bishop, 2003) and were then retested on the BPVS. Correlations between TROG scores and the other measures are available in the supplement (Figure S1, *§3);* here we focus on vocabulary as this was tested twice. Note that the raw BPVS and TROG scores could not be converted to standardized scores due to many children in our sample being below the minimum age in the norming samples (3 years and 4 years, respectively).

Eye-tracking Task. In this visual-world task, children listened to sentences while viewing three pictures on a screen, each of which depicted a potential final word (Table 1 and Figure 1). We created 15 sets of items, i.e., sets of three pictures with three associated sentences. For each set, we created two different predictive sentences and a non-predictive sentence. We had two different predictive sentences to control for potential differences in salience between the pictures - one of the predictive sentences made one of the pictures highly predictable and a different one implausible, while the other predictive sentence made the latter picture highly predictable and the former implausible; the third picture was always mildly predictable. To illustrate, for the following set of pictures - A. bone, B. slippers, C. pyjamas - the predictive sentence *Alfie's dog likes to chew on the…* induced the graded ordering A>B>C, while the other predictive sentence *When you go to bed, you wear…* induced the opposite ordering, C>B>A; the non-predictive sentence was *Now, Craig is looking for the …*, inducing the ordering A=B=C. We refer to these three sentence conditions as A-biasing, C-biasing, and Neutral. Importantly, we developed the items through pretesting with adults, and then confirmed the graded predictability pattern through a pre-test with 24 preschoolers: Children listened to sentence contexts (i.e., sentences without the final word as in the examples above), and then the experimenter asked them for help "finishing off the story"; they chose the picture they thought was the best end for the story, and then the procedure was repeated with the remaining two pictures, so that they implicitly ranked the pictures from best to worst completion (see *§2* in Supplementary Materials for further details). On average, after A-biasing sentence contexts, children chose the pictures in the order A>B>C 76% of the time, range [62.5%,87,5%]; after C-biasing contexts, the pictures were chosen in the order C>B>A 73% of the time, range [62.5%, 100%]; finally, after neutral contexts the average proportion of children who converged on the most preferred ordering (which differed across sentences) was much lower, at 45%, range [37.5%,75%].

We also varied which picture was eventually named. Following predictive A-biasing and C-biasing contexts, children heard either the predictable word (i.e., A or C, e.g., *When you go to bed, you wear pyjamas*) or the mildly predictable word (i.e., B … *wear slippers*; counterbalanced across lists); the unpredictable picture was never named. Neutral contexts could be followed by either A, B or C.

Participants completed two blocks of 15 trials, such that they encountered each item set once per block, with items always assigned to different conditions in each block, counter-balanced across six lists. Participants heard 5 A-biasing, 5 C-biasing, and 5 neutral trials in each block, so they heard twice as many predictive sentences as neutral sentences. Note that, because neutral sentence contexts followed by B were particularly critical for our analyses (as they were compared to predictive contexts followed by B), these trials were always placed in the first block, so that participants were more likely to complete them. Neutral contexts followed by A or C occurred in Block 2.

Each trial began with a 2-second silent preview of the objects, after which participants heard the sentence, followed, two seconds later, by an instruction to point to the object mentioned in the sentence. The experimenter then noted the child's response, triggered a "reward" screen (a cartoon image plus a cheery sound), and began the next trial. Trial order within blocks was randomized by participant, and object positions were counterbalanced across trials. Audio stimuli were recorded by a male Scottish English speaker, and images were sourced online and scaled to 300x300px.

A REDn Scientific eye-tracker (SensoMotoric Instruments GmbH, www.smivision.com) tracked both eyes at 30Hz. We performed calibration before each block using a 5-point grid. Only right-eye data (left for one child, who had impaired right-eye vision) were analyzed.

Data Analysis and Results

Our first set of analyses focused on the cross-sectional data from all 215 children who took part in Phase 1 (*Cross-sectional analyses*). We first conducted group-level analyses using data from the eye-tracking task to assess whether children were able to generate graded predictions (*The development of graded predictions*) and took longer to process a word when it disconfirmed a prediction than when no prediction was disconfirmed (*The development of revision skills*). The power of these analyses, which used linear mixed-effects models, depends both on sample size and the number of trials per condition (Brysbaert & Stevens, 2018); while our design was novel and not directly comparable to any published studies, our sample size was considerably larger than previous eye-tracking studies of linguistic prediction in children (see *Participants* above) and the number of trials per condition (10) was comparable (6 in Gambi et al., 2016; 10 in Gambi et al., 2018; 10 in Mani et al., 2016; 12 in Mani and Huettig, 2012; 16 in Borovsky et al., 2012). These group-level analyses were followed up with individual difference analyses: We assessed how each child's concurrent language skills (i.e., receptive vocabulary) was related to their ability to generate graded predictions (*The development of graded predictions*), their ability to revise after having a prediction disconfirmed (*The development of revision skills*), and their word processing speed following neutral contexts that do not elicit prediction (*The development of processing speed*). Post-hoc sensitivity analyses showed that, with a sample size of 215, we had 95% power to detect a relation with |rho| = 0.240 (correlation) or $f^2 = 0.061$ (multiple regression); that is a small effect size.

Our second set of analyses was conducted on the sub-sample of children (N=55) whose vocabulary was tested twice, to assess whether these same language processing abilities measured in Phase 1 using eye-tracking explain unique variance in vocabulary development between Phase 1 and Phase 2 (*Longitudinal analyses*). These analyses were exploratory. Post-hoc sensitivity analyses analogous to the ones conducted for Phase 1 showed that, with a sample size of 55, we had 95% power to detect a relation with |rho| = 0.444 (correlation) or $f^2 = 0.245$ (multiple regression); that is a medium effect size, though it should be noted that the true power may be lower than suggested by these sensitivity analyses because of measurement error (Williams, Zimmerman, & Zumbi, 1995).

All analyses were performed in R (Version 3.13) using functions *lme4* (Bates, Maechler, Bolker, & Walker, 2015) and *lm*. Nominal alpha was set to .05 in all analyses. Key analyses used a regression approach to simultaneously test all core hypotheses and take into account relevant control variables, thus limiting alpha inflation due to multiple comparisons.

Before analysis, the eye-tracking data was pre-processed to assign fixations to areas and time windows of interest. We drew 300x300px areas of interests (AOIs) around each picture, and analyzed fixations to these AOIs in 100ms-bins. Fixations outside the AOIs were excluded from analysis. Analyses focused on two time-windows: a *prediction window* lasting from 1000ms before the final word onset to 100ms after (to account for the time it takes to launch a saccade; Trueswell, 2008); and a *recognition window*, from 100ms after final word onset to 1000ms after its offset. Thus, the prediction window had constant duration (1100ms) but its onset was variable relative to sentence onset, as the onset of the final word occurred at a variable position (M = 2179ms from sentence onset, range [1190ms, 4148ms]); in contrast, the duration of the recognition window was variable (M = 1541ms, range [1317ms, 1856ms]), as final words varied in length. We discarded trials on which children's pointing or speech overlapped with the sentence (4.6%), as well as trials on which no gaze data was recorded for more than 40% of the duration of the time window of interest (prediction: 6.05%; recognition: 4.38%). The prediction window was used to assess whether children's predictions are graded (*The development of graded predictions*), and the recognition window was used to assess children's word processing skill (*The development of processing speed*). Both windows were used to assess children's revision skill (*The development of revision skills*), as we describe below.

Cross-sectional analyses.

The development of graded predictions. If children's predictions are graded then, as a predictive context unfolds, looks to the predictable picture should become more likely than looks to the mildly predictable picture, which in turn should become more likely than looks to the unpredictable picture. Figures 2A and 2B show how this behavior emerges, for both A-biasing contexts (left panels) and C-biasing contexts (middle panels, neutral contexts are shown in right panels). Figure 2A splits the data by age, and Figure 2B by raw vocabulary size.

To statistically analyze how the pattern of gaze evolves over time from the beginning to the end of the prediction window, we applied Growth Curve Modelling (Mirman, 2014;

note that these growth curves thus model change over the sentence, not longitudinal change over age). We began by calculating difference curves that compared gaze during predictive contexts to gaze during neutral contexts (see Figure 2C). This difference curve approach is necessary because comparing looks across pictures within a condition would violate independence assumptions (see Kukona, Fang, Aicher, Chen, & Magnuson, 2011), since the eyes can only fixate on one picture at a time; instead, we compare how the difference in proportion of looks between conditions (predictive vs. neutral contexts) varies across the three pictures. We applied an empirical logit (elog) transformation (Barr, 2008) to the proportion of looks to each picture before computing the difference curves, thus the y axis in Figure 2C represents the empirical log odds of gazing at each picture in the predictive contexts compared to the neutral contexts. For confirmation that age and vocabulary effects are also seen in the difference curves, see Figure S2, *§4.1,* Supplement).

Recall from the Methods section that each set of pictures was paired with two predictive sentences, A-biasing and C-biasing, to control for baseline salience differences across pictures. At the analysis stage, we collapsed across these conditions to increase power, so we will describe the findings in terms of looks to Predictable pictures (i.e., A pictures following an A-biasing context and C pictures following a C-biasing context), Unpredictable pictures (i.e., C pictures following an A-biasing context and A pictures following a C-biasing context), and Mildly Predictable pictures (i.e., B pictures; see *§4.2* in the Supplement for confirmation that the pattern held for each type of predictive sentence). Our growth curve regressions quantified the gradedness of children's predictions across the three pictures using two dummy-coded contrasts, one capturing the preference for Predictable vs. Mildly predictable pictures, and the other the dis-preference for Unpredictable vs. Mildly predictable pictures.

We used orthogonal polynomials to model how these preferences for the pictures changed over the course of the prediction window; a linear time term (*time*) modelled overall increases or decreases in preference, while a quadratic term ($time^2$) modelled differences in curvature, with larger absolute values indicating a steeper change in looks over time. To capture how children's graded predictions emerged as the sentence unfolded, we included interactions between the two dummy contrasts and the two time terms. The model also included age and linguistic knowledge (raw vocabulary size) as (centered) covariates, and their interactions with all other terms, so that the lower-order predictors would reflect performance of a child of average age and linguistic knowledge in our sample. Thus, the final model had the form, in lmer syntax, elog(Prop. Predictive) – elog(Prop. neutral) ~ 1 + (time+time^2)*(Predictable-Mildly predictable + Unpredictable-Mildly predictable)*(Age+Vocabulary), plus maximal by-participant random effects. Note that we only report a by-participant analysis (i.e., collapsing over items to yield more robust estimates and aid convergence), but the by-items analysis was consistent (see *§4.3* in the Supplement).

Table 2 shows the results of the model, excluding the age/vocabulary effects and their interactions, which are reported in the supplement (Table S5, *§4.4*). The model confirmed the pattern of graded predictions in Figure 2C. Preschoolers showed an overall preference for predictable over mildly predictable pictures (*intercept*, t=8.82), and also a dis-preference for unpredictable pictures compared to mildly-predictable pictures (*intercept*, t=-2.05). Over the analyzed window, the preference for predictable pictures was quite stable (*time*, t = 1.70), showing only a slight but significant tendency to level off towards the end of the window ($time^2$, t = -2.01). In contrast, the dis-preference for unpredictable compared to mildly-predictable pictures became more pronounced with time (*time*, t=-2.99), particularly towards the end of the window (*time*^{2}, t=-3.24). In sum, we found clear evidence for graded predictions in our sample of 2-to-5-year-olds.

INSERT TABLE 2 HERE

While Table 2 shows the estimated behavior of the average child in our sample, Figures 2A and 2B suggest that there are also interesting age and vocabulary-related differences in children's ability to generate graded predictions. Thus, we next explored how graded predictions varied across age and raw receptive vocabulary size. While the growth-curve model fitted above includes age and vocabulary effects and their interactions with the parameters reported in Table 2 (see *§4.4* of the Supplement), it is not ideally suited to address this question because it models the preference for predictable pictures separately from the dispreference for unpredictable pictures (i.e., as two different parameters). In order to capture individual differences in the overall gradedness of children's predictions, we instead computed a combined graded prediction measure, capturing both the preference for the most predictable continuation and the dispreference for the unpredictable continuation, and then we examined the relation between children's linguistic knowledge and this combined measure.

To compute this combined measure, we analyzed raw gaze proportions averaged over the last 400ms of the prediction window. We chose this shorter window because, based on visual inspection of Figure 2, the overall size of the prediction effect was largest here. For each participant, we first subtracted the mean gaze proportion for each type of picture during a neutral context from the mean gaze proportion for the same type of picture during a predictive context. We then used these difference scores to compute the mean preference for predictable over mildly predictable pictures (i.e., mean gaze proportion to predictable pictures minus mildly predictable pictures averaged over the last 400ms of the prediction window) and the mean dis-preference for unpredictable pictures (mean gaze proportion to unpredictable minus mildly predictable pictures averaged over the same time window). The combined measure of graded prediction skill was then defined as the mean preference minus the mean dis-preference. This combined measure was correlated with both age ($r(123) = .369$, $p < .001$) and vocabulary ($r(123) = .326$, $p < .001$; see Figure 4A). Importantly, incorporating the gradedness of prediction appeared to increase the strength of this relation: When age and vocabulary were each separately correlated with the two individual components of the graded prediction measure (i.e., the preference for predictable picture and the dispreference for unpredictable pictures), then the relevant associations were weaker or indeed non-significant ($r < .22$; see *§4.5* of the Supplement). Thus, this suggests that measuring the gradedness of predictions captured an important component of children's developing language processing skills.

Finally, we looked to see if there was a relation between children's prediction ability (via the combined prediction measure above) and their linguistic knowledge, i.e., vocabulary size, over-and-above differences that are associated with getting older. We compared the relative fit of a linear model regressing graded prediction score against age, to the fit of a model that additionally incorporated children's vocabulary score (using an F test to compare the residual sum of squares of the two models); the fit of the latter model should be significantly higher if vocabulary explains additional variance, above-and-beyond age. However, this was not the case ($F(1, 212)=0.599$, $p>.250$), suggesting that, while children's graded prediction ability may be a better indicator of their linguistic knowledge compared to their ability to anticipate the most predictable continuation or to rule out implausible continuations, this relation may yet be fully explained by other skills that also improve with age.

INSERT FIGURE 2 HERE

The development of revision skills. Our first set of analyses showed that children's ability to differentiate between multiple predictable continuations grows with age and vocabulary knowledge. But while this suggests that children's predictions become more sophisticated as they develop, it also raises the question of how the complementary ability to revise (inaccurate) predictions develops. To address this question, we first conducted group-level analyses to test whether recognition is indeed slower, in children, following a disconfirmed prediction than when no prediction is disconfirmed. We then assessed how a measure of revision skill ("predict-and-redirect", after Reuter et al., 2019) relates to age and vocabulary.

To test the proposal that (inaccurate) predictions hinder processing, we analyzed the speed with which children recognized the mildly-predictable picture after predictive versus neutral contexts. The key idea here is that the neutral context provides a baseline measure of how quickly children can recognize the spoken name of the mildly-predictable picture when other pictures are equally expected (for confirmation that looks to mildly-predictable B pictures are roughly as likely as looks to the other two pictures after a neutral context, see Figures 2A and 2B, right panels). However, after a predictive context the predictable picture is significantly more expected than the mildly predictable picture (as shown in *The development of graded predictions*). Thus, if the mildly-predictable picture is named instead of the predictable picture, we may see a delay in recognizing its name following a predictive context compared to the neutral context. We thus analyzed the time (in milliseconds) that it took children to gaze at the mildly predictable (B) picture, across predictive and neutral contexts (Context, contrast-coded and centered) on trials on which participants were not already gazing at that picture at 100ms following name onset (cf. Barr, 2016; Fernald, Zangl, Portillo, & Marchman, 2008); the median number of trials contributed to this analysis by each child was 3 in both the neutral and the predictive condition (out of 5 possible trials in each condition). Our model had the structure Latency ~ 1 + Context *(Age + Vocabulary), plus maximal random effects by item, and random intercepts by participants (by-participant slopes for Context were estimated to be close to zero and dropped for convergence).

We found strong evidence that inaccurate predictions hinder processing. Overall, children took longer to orient their attention towards the mildly predictable (B) picture after this picture was named following a predictive context compared to a neutral context (Figure 3C), indicating that having predicted a different picture, and having that prediction disconfirmed, slowed down recognition ($B = -95.51$, $SE = 25.28$, $t = -3.78$, CI = [-145.06,-45.96]); the full model is available in *§5* of the Supplement, Table S6). Thus, the average child in our sample experienced costs when having a prediction disconfirmed. Moreover, as Figures 3A and 3B suggest, the magnitude of this cost was positively associated with both age and vocabulary size (i.e., there were significant interactions between Context and Age, and Context and Vocabulary, both t's > 2.6; see Tables S7 and S8 in *§5* of the Supplement for full model summaries).

Next we examined the development of revision skills: Given that children experience costs associated with making inaccurate predictions, the ability to efficiently revise following the encounter with an unexpected word should be critical. To characterize revision skill, we computed a “predict-and-redirect” measure (Reuter et al., 2019), which captured how children responded when a predictive context was followed by a mention of the mildly predictable picture. We subtracted mean proportion gaze to the mildly predictable picture during the last 400 ms of the prediction window from mean proportion during the recognition window (after Reuter et al., 2019; we could not compute this measure for two participants due to missing data). Thus, a higher score on the measure indicates that the child initially gazed to the most predictable image, but subsequently quickly redirected their attention when those predictions were disconfirmed. Importantly, we found that revision skill was strongly correlated with both age (r(211)=.423, p<.001) and vocabulary (r(211) =.493, p<.001; see Figure 4B). Moreover, and unlike skill at prediction on its own, we found an association with vocabulary over-and-above the effect of age (F(1,210)=18.235, p<.001; when comparing a linear regression model including age and vocabulary to a model including age only). Thus, these data suggest a unique relation between children's current linguistic competence and their ability to rapidly predict-and-revise, which cannot be explained away by other factors that improve with age.

INSERT FIGURE 3 HERE

The development of processing speed. Finally, to measure how quickly children recognize spoken words, we followed previous work (Fernald & Marchman, 2012; Fernald et al., 2006; Marchman & Fernald, 2008), and used the average time (in milliseconds) of the first fixation to the named picture during the recognition window. To compute this measure, we used only data from neutral sentences, so we could assess children's general word processing ability in the absence of strong contextual support for prediction. Following standard practice, we included only trials on which participants were not already gazing at that picture at 100ms following name onset (cf. Barr, 2016; Fernald, Zangl, Portillo, & Marchman, 2008). Confirming previous reports (Fernald & Marchman, 2012; Fernald et al., 2006; Marchman & Fernald, 2008), children's word processing speed increased with age (r(213=-.297, p<.001) and vocabulary (r(213)=-.294, p<.001; see Figure 4C). Somewhat surprisingly, however, vocabulary did not significantly explain any unique variation in processing speed over-and-above the effect of age (F(1, 212) = 2.078, p =.151; when comparing a linear regression model including age and vocabulary to a model including age only).

Summary of cross-sectional analyses. In sum, in our large sample of 2- to 5-year-olds, we found that three different measures of children's language processing ability – of graded prediction skill, of revision skill, and of processing speed – increase with age and vocabulary knowledge. Of the three measures, only revision skill was associated with vocabulary over-and-above the effect of age, and appears therefore to have the strongest link to children's concurrent structural knowledge of language. However, cross-sectional analyses cannot address the question of how prediction, revision, and processing speed are associated with later language development. To provide a preliminary answer to that question, we turned to the longitudinal data.

INSERT FIGURE 4 HERE

Longitudinal analyses. In these exploratory analyses, we assessed how prediction, revision and processing speed were associated with changes in vocabulary size from Phase 1 to Phase 2 (see Supplement, *§6*, for plots showing that age and vocabulary distributions at Phase 1 were similar across the full sample and longitudinal sub-sample). The three skills were quantified using the same summary statistics as in the cross-sectional analyses. We captured prediction skill using the combined measure – i.e., through children's preference for predictable pictures minus the dispreference for unpredictable pictures (see Supplement, *§7*, for evidence that neither the preference nor the dispreference measure alone were strongly predictive of changes in vocabulary size); we captured revision skill thought the "predict-and-redirect" measure (Reuter et al., 2019), and finally we captured processing speed using the average timing of the first fixation to the named picture during the recognition window (e.g., Fernald et al., 2006).

Note that, because recruitment in Phase 2 was opportunistic, our sample was highly variable: It contained children from a wide range of ages who, furthermore, were retested at different intervals. Recognizing that the nature of our sample made a simple comparison between raw vocabulary scores at Phase 2 and raw vocabulary scores at Phase 1 inappropriate, we endeavored to control for some of this variability post-hoc during analyses. Specifically, analyses that do not control for the child's age at the time they were first tested (in Phase 1) and the duration of the test-retest interval could confound interesting individual differences in the rate of vocabulary development with group-level (i.e., average) differences in the rate of vocabulary development across age groups. Thus, we needed a measure of children's vocabulary knowledge that would take into account the average vocabulary size of their age cohort, and would hence be informative about whether the child's vocabulary grew faster or slower than would typically be expected between Phase 1 and Phase 2.

We derived a measure with these properties as follows. Since we could not work with standardized scores (these were not available for children below 3) we instead converted raw BPVS scores into equivalent linguistic ages for all children in our longitudinal sub-sample. Linguistic age is defined as the age of the average child with the same raw BPVS score in the BPVS-II norms. Thus, comparing linguistic age to chronological age provides an indication of whether a child is more or less linguistically advanced than the average child in the BPVS-II norms, and so we focused on this relative measure. Specifically, we expressed linguistic age as a percentage increment of chronological age; e.g., for a 36-month-old child with a linguistic age of 42 months during Phase 1, their linguistic age would be (42-36)*100/36 = 16.7% higher than their chronological age, indicating that they are more advanced linguistically than the average child. If this child were retested 6 months later (chronological age: 42 months) and found to have a linguistic age of 49 months at Phase 2, this would mean their linguistic age would still be (49-42)*100/42 = 16.7% higher than their chronological age; that is, over the test-retest interval, the child's vocabulary would have grown at the same speed as the that of the average child. But if the same child's linguistic age at 42 months were instead 54 months, the child's linguistic age would have increased to be (54-42)*100/42 = 28.6% higher than their chronological age by the end of the test-retest interval. In other words, this would suggest the child's vocabulary grew faster than that of the average child between Phase 1 and Phase 2, and specifically that their rate of vocabulary development was 28.6%-16.7% = 11.9% higher than that of the average child.

Importantly, having defined the rate of vocabulary change as the difference between linguistic age expressed as a percentage increment of chronological age at Phase 2 and Phase 1, we could directly compare children who were retested at different intervals, because this measure uses the performance of the average child in BPVS-II norms as a reference point. Using our measure of vocabulary change, one child's score was exceptionally large (>200%), so it was discarded, leaving N = 54. After removing this child, the average rate of vocabulary change was -3.41% (i.e., not different from zero: $t(53) = -1.03$, $p = 0.31$). However, there was still considerable variation in the sample, range [-67.93%, +53.38%], suggesting it made sense to ask whether any of that variation was related to children's processing skills at Phase 1. A negative score here means that the child's vocabulary grew less rapidly than expected based on BPVS-II norms, whereas a positive score means that the child's vocabulary grew faster than the average child's; a score of zero means the child's vocabulary grew at the same rate as the average child's (see Supplement, *§9*, Table S9, for a table reporting each child's rate of vocabulary change).

In sum, our measure captures more than just absolute increases in the size of children's vocabulary – it captures the degree to which a child's vocabulary is growing faster or slower than their peers. It thus makes it possible to ask whether children who learnt vocabulary at faster-than-average rates between Phase 1 and 2 are those whose processing skills (graded prediction, revision, processing speed) were more advanced in Phase 1. To answer this, we first used separate linear regressions to assess the contribution of each processing skill, and then followed these up with a multiple regression analysis to establish whether any of the processing skills explained variance in children's rate of vocabulary change over-and-above the others. The processing measures were all converted to z scores to facilitate comparison of their effect sizes. Even though raw vocabulary in Phase 1 did not correlate with rate of vocabulary change, $r(52) = -.08$, $p > .250$, we additionally controlled for this variable (centered) in all analyses, to capture any residual differences in the rate of vocabulary change across different stages of linguistic development. (The correlation between rate of vocabulary change and age at Phase 1 was somewhat higher, $r(52) = .13$, $p > .250$, but additional analyses controlling for age at Phase 1, instead of raw vocabulary at Phase 1, yielded consistent findings; see Supplement, *§8*).

Previous work has found that vocabulary grows faster in children who recognize spoken words more quickly (Fernald et al., 2006), and we replicated that result here, showing that children with faster processing speed at Phase 1 were more likely to grow their vocabulary at faster-than-average rates between Phase 1 and Phase 2 ($B = -7.16$, $SE=3.33$, $t= -2.15$, $p = .036$, see Figure 5A). Next, we asked whether a similar relation was also found for our measures of prediction and revision skill. Interestingly, children with stronger skills at graded prediction also grew their vocabulary at faster-than-average rates ($B = 6.69$, $SE= 3.28$, $t=2.04$, $p =.047$; Figure 5B), although the relevant statistical comparison only just reached significance. However, children with stronger revision skill did not show significant evidence of faster-than-average improvement in vocabulary knowledge over time ($B = 3.13$, $SE = 3.69$, $t = 0.85$, $p>.250$; Figure 5C).

INSERT FIGURE 5 HERE

These results confirm previous reports that inter-individual variation in the ability to rapidly recognize spoken words explains inter-individual variation in the speed of vocabulary development (Fernald et al., 2006), and suggest that the ability to form graded expectations about upcoming words may also play a similar role. In contrast, the ability to efficiently revise inaccurate expectations did not appear to explain inter-individual variation in the speed of vocabulary development, despite being associated with concurrent linguistic knowledge (see *The development of revision skills*). Thus, we dropped revision skills from further analyses, and instead focused on assessing whether prediction skill and processing speed are independent contributors to the rate of vocabulary change.

To do so, we entered both measures into a multiple regression (again, controlling for vocabulary in Phase 1, centered). Neither measure individually was now a reliable predictor: Graded prediction, B = 5.35, SE = 3.31, t = 1.62, p = .112; Processing speed, B = -5.90, SE = 3.36, t = -1.75, p = .086, suggesting that some of the variation in the rate of vocabulary change explained by each of the two processing skills is also explained by the other – that is, the two processing skills explain overlapping variance in the rate of vocabulary development. Indeed, this was confirmed in a commonality analysis (Ray-Mukherjee, Nimon, Mukherjee, Morris, Slotow, & Hamer, 2014), performed using the R package *yhat* (Nimon, Oswald, & Roberts. 2016): According to this, of the total variance explained by the multiple regression model (R^2 = .135), processing speed accounts uniquely for 39.38%, graded prediction skill accounts uniquely for a comparable 33.53%, and together they account for a further 21.75%.

A potential interpretation of this result is that these two abilities – prediction skill and processing speed – both influence linguistic development via a common mechanism; in particular, both could be considered as distinct measures of a single underlying ability to fluently process language. Consistent with this, we found that the rate of vocabulary change was predicted by a combined measure, corresponding to the sum of the two scores (with processing speed sign-reversed, so higher values correspond to faster recognition). Specifically, a linear regression model containing the combined measure (and again controlling for raw vocabulary in Phase 1) explained a small but significant amount of variance in the rate of vocabulary change (R^2 = .135, F (2,51) = 3.98, p = .025), and model comparison (using an F test to compare the models' residual sum of squares) showed that including this combined measure significantly improved the fit of the model compared to a baseline model only including raw vocabulary at Phase 1 (B = 8.82, SE = 3.21, F(1,51) = 7.53, p =.008).

In sum, our longitudinal analyses provide preliminary evidence that prediction skills may play a facilitatory role in children's language development, in a similar manner to how word recognition speed does. These analyses also highlight the intriguing possibility that both prediction and processing speed may contribute to vocabulary acquisition through enhancing children's fluency at processing language.

Discussion

Using a sensitive eye-tracking task, we investigated the relation between vocabulary acquisition and language processing in a large sample of pre-schoolers. In particular, we examined how children's vocabulary knowledge relates to three processing skills: the ability to generate graded predictions, the ability to recover from incorrect predictions, and the ability to recognize spoken words. We then followed up a subset of the children to further explore how processing skills relate to inter-individual variation in how rapidly vocabulary grows over time.

Our study revealed important developments in children's sentence processing skills, and how these skills relate to concurrent linguistic knowledge; it also provided some preliminary evidence regarding the relation between processing skills and the rate of subsequent language development. First, between the ages of 2 and 5, children's predictions become increasingly sophisticated, as they become more sensitive to graded distinctions in predictability. However, we also found that as prediction skills emerge over the preschool years, so do the costs associated with recognizing a word when another, more likely word has (incorrectly) been predicted in its place. Second, all the language processing skills that we examined – the abilities to make graded predictions, to revise incorrect predictions, and to recognize words fluently – were associated with concurrent vocabulary size, but only the ability to revise incorrect predictions was related to concurrent vocabulary knowledge over-and-above the effect of age. Third, we found preliminary evidence that the degree to which children show graded sensitivity when generating linguistic expectations may be associated with the rate at which their vocabulary will grow over following months. Similarly, we replicated previous reports that children's ability to quickly recognize a spoken word is related to how rapidly their vocabulary knowledge will grow (Fernald et al., 2006). In contrast, children's skill at revision was not related to inter-individual variation in the rate of vocabulary development in our longitudinal sample. Moreover, children's graded prediction skills and their word recognition skills were not independently related to the rate of vocabulary change; rather, much of the inter-individual variation explained by each of these predictors was also explained by the other. Below, we begin by discussing how the first set of findings adds to our knowledge of children's sentence processing skills; we then consider the second and third set of findings– on cross-sectional and longitudinal associations (respectively) between processing skills and vocabulary knowledge –and assess how they can constrain hypotheses about the relation between children's in-the-moment processing of linguistic input and the development of linguistic knowledge.

First, our data provide a clearer picture of how children's language processing skills develop in the preschool years. The finding that preschoolers consider multiple alternatives in parallel, each proportionally to its predictability in context, adds to previous evidence for a high degree of sophistication in preschoolers' linguistic predictions (Borovsky et al., 2012;

Gambi et al., 2016; Havron et al., 2019; Lindsay, Gambi, & Rabagliati, 2019; Mani & Huettig, 2012; Mani et al., 2016). Previous findings had already shown that preschoolers use their knowledge of semantics (e.g., Borovsky et al., 2012) and linguistic structure (e.g., Gambi et al., 2016) when they generate predictions about the single most likely continuation for a transitive sentence, and that their predictions are sensitive to the strength of the semantic association between a word and the sentence context (Mani et al., 2016). However, to our knowledge the current study is the first to directly show that preschoolers are sensitive to graded distinctions in predictability - i.e., that they distinguish not only between more predictable and less predictable words, but also between less likely words and completely implausible words. This is important because gradedness is a key feature of adult linguistic predictions (e.g., Staub et al., 2015).

We also showed that preschoolers experience a slow-down in word recognition when they encounter a word that is comparatively unexpected. This finding has important implications for our understanding of the relation between prediction, processing speed, and language development. Previous work has shown that recognition of a word is facilitated when it occurs in a predictive context (e.g., Lew-Williams & Fernald, 2007), but our finding shows that predictive contexts can be a double-edged sword, slowing the recognition of plausible but less-likely words. Importantly, this finding held under quite stringent conditions. In particular, recognition of a moderately predictable word was slowed down if an alternative word was much more predictable, as compared to a neutral baseline where the same word was moderately predictable, but no other word was strongly predictable. This shows that there are potential disadvantages for children who continuously generate predictions as they process sentences, particularly if their language model is likely to be inaccurate (and thus generates many incorrect predictions; Omaki & Lidz, 2015).

Our second and third set of findings concern the cross-sectional and longitudinal relation between children's language processing skills and their vocabulary knowledge. Our eye-tracking task allowed us to derive three different measures of children's skill at processing language - graded prediction, revision, and processing speed, and we will consider each in turn. Starting with prediction skill, while previous studies reported positive associations between children's ability to predict and their concurrent vocabulary knowledge (Borovsky et al., 2012, Mani & Huettig, 2012, Mani et al., 2016) our study is the first to suggest that the degree to which children's predictions are graded may capture important variation in the speed of their linguistic development. Interestingly, the concurrent association between graded prediction skill and vocabulary knowledge in the present study could be explained by age-related changes in the ability to generate graded predictions (see also Gambi et al., 2016; Gambi et al., 2018), suggesting that this relation may be explained by other underlying skills that improve with age, such as domain-general processing speed. However, our longitudinal analysis did suggest that graded prediction skill may contribute to inter-individual variation in the speed with which vocabulary grows over time, perhaps as one component of a broader processing-speed factor (see below). With the caveat that this preliminary finding requires replication, it does suggest that prediction skills can act to facilitate language development. In addition, our data clearly show that the strongest relation between concurrent vocabulary size and prediction skill was for the measure that incorporated gradedness, i.e., the measure that accounted for both the preference for predictable pictures and the dispreference for unpredictable pictures. Thus, our data suggest that taking into account the degree of gradedness of children's linguistic predictions may be important for fully characterizing the relation between prediction during language processing and language knowledge. We suggest that it will be important for future longitudinal studies to incorporate a measure of graded prediction skill.

Our findings also shed light on the relation between revision skill and vocabulary development. Cross-sectionally, we found that those children who are more efficient at revising a strong but incorrect prediction are also more linguistically advanced than their peers, which is consistent with recent work by Reuter et al. (2019), who found that children with stronger revision skills were better at learning the meanings of new words that were encountered in contexts that required revision. However, the interpretation of that finding was unclear: do stronger revision skills make children better learners, or do more advanced linguistic and word-learning skills allow children to engage in more accurate processes of revision (cf. Rabagliati et al., 2015)? Our longitudinal data may help inform a preliminary answer to this question. If the process of linguistic revision is a key driver of learning, then we would also expect revision-related processing skills to explain unique variance in the rate of vocabulary change over time, and not just in concurrent linguistic skills. However, we found no evidence for this in our longitudinal sample, providing no clear indication that a predict-and-revise mechanism drives language development. Thus, we suggest that the strong cross-sectional relation between revision skill and vocabulary knowledge may result from changes in linguistic knowledge that drive changes in revision processing skills, rather than the other way around. Importantly, however, since our longitudinal analyses were exploratory, more research (using less heterogenous longitudinal samples) will be needed to confirm this suggestion.

In contrast, we confirmed previous findings that processing speed is linked to the speed of language development, as children who were faster to recognize words also had a faster rate of vocabulary growth over the next few months (Fernald et al., 2006; see also Peter et al., 2019). Further, our analyses suggested that the positive relation between processing speed and the speed of linguistic development overlaps with that of prediction skill: To the extent that children’s skill at graded prediction explains variance in the rate of vocabulary change, this explained variance is importantly shared with processing speed. We suggest that this finding is consistent with the hypothesis that both skills may benefit language development via the same mechanism: Prediction and processing speed may contribute overlapping variance to vocabulary change over time because they both enhance children's fluent language comprehension. In particular, children who can extract meaning more quickly from sentence contexts, either via faster bottom-up processing of the input (processing speed) or via prediction of the input (prediction skill), are at an advantage when it comes to tasks such as making inferences about the meaning of unknown words (Fernald et al., 2008). We further speculate that this facilitatory effect of prediction on fluent language comprehension may on the whole outweigh the fluency costs associated with incorrect predictions.

In sum, we suggest that our findings are overall most consistent with models of linguistic development in which both prediction and processing speed benefit language development thanks to the facilitative effect they have on fluent processing of linguistic input. By facilitating fluent language processing, both skills contribute to freeing up resources during online processing of sentences, which can be dedicated to other tasks, including encoding the form of unknown words into memory, and inferring the meaning of those words from their linguistic and non-linguistic context.

Conclusion. Our study provides a first step towards better understanding the link between prediction and language development. We showed that graded predictions about upcoming words become more sophisticated between the ages of 2 and 5, and found suggestive evidence for a relation between children's skill at generating graded predictions and their subsequent rate of linguistic development. At the same time, we also replicated the relation between processing speed and inter-individual variation in the speed of language development, and found that some indication that these two processing skills – prediction and fluent word recognition – may explain overlapping variance in the rate of linguistic development. Thus, we suggest that graded prediction ability may support linguistic development by increasing the fluency with which children process language.

Acknowledgements

We would like to thank all participating children, families and nurseries, as well as Alexander Robertson for help with recruiting participants. This project was supported by a Leverhulme Trust Research Project Grant (RPG-2014-253) to HR and MP and an ESRC Future Research Leaders award (ES/L01064X/1) to HR.

References

Aravind, A., de Villiers, J., Pace, A., Valentine, H., Golinkoff, R., Hirsh-Pasek, K., ... & Wilson, M. S. (2018). Fast mapping word meanings across trials: Young children forget all but their first guess. *Cognition, 177,* 177-188.

Baayen, R. H., Davidson, D. J., & Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59,* 390-412.

Barr, D. J. (2016). Visual world studies of conversational perspective taking: Similar findings, diverging interpretations. In P. Pyykkönen-Klauck, P. Knoeferle, & M. Crocker (Eds.), *Visually situated language comprehension* (pp. 261-290). Amsterdam: John Benjamins.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67,* 1-48, doi:10.18637/jss.v067.i01.

Bishop, D. (2003). *Test for Reception of Grammar, TROG-2*: Pearson.

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*, 417-436.

Brysbaert, M., & Stevens, M. (2018). Power Analysis and Effect Size in Mixed Effects Models: A Tutorial. *Journal of Cognition, 1*, 9, doi:10.5334/joc.10

Carter, B. T., Foster, B., Muncy, N. M., & Luke, S. G. (2019). Linguistic networks associated with lexical, semantic and syntactic predictability in reading: A fixation-related fMRI study. *NeuroImage, 189*, 224-240.

Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review, 113*, 234-272.

Choi, Y., & Trueswell, J. C. (2010). Children's (in) ability to recover from garden paths in a verb-final language: Evidence for developing control in sentence processing. *Journal of Experimental Child Psychology, 106*, 41-61.

Duff, F. J., Reen, G., Plunkett, K., & Nation, K. (2015). Do infant vocabulary skills predict school-age language and literacy outcomes? *Journal of Child Psychology and Psychiatry, 56,* 848 – 856.

Dunn, L., Dunn, L., Whetton, C., & Burley, J. (1997). *The British Picture Vocabulary Scale—Second Edition*. Windsor: NFER-Nelson.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., . . . Stiles, J. (1994). Variability in early communicative development. *Monographs of the society for research in child development, 59*, 1-173.

Fernald, A., & Marchman, V. A. (2012). Individual differences in lexical processing at 18 months predict vocabulary growth in typically developing and late-talking toddlers. *Child Development, 83*, 203-222.

Fernald, A., Marchman, V. A., & Hurtado, N. (2008). *Input affects uptake: How early language experience influences processing efficiency and vocabulary learning.* Paper presented at the 7th IEEE International Conference onDevelopment and Learning.

Fernald, A., Perfors, A., & Marchman, V. A. (2006). Picking up speed in understanding: Speech processing efficiency and vocabulary growth across the 2nd year. *Developmental psychology, 42*, 98-116.

Fernald, A., Zangl, R., Portillo, A. L., & Marchman, V. A. (2008). Looking while listening: Using eye movements to monitor spoken language. In I. A. Sekerina, E. M. Fernandez, & H. Clahsen (Eds.), *Developmental psycholinguistics: On-line methods in children's language processing* (Vol. 44, pp. 97-135). Amsterdam/Philadelphia: John Benjamins.

Friend, M., Smolak, E., Liu, Y., Poulin-Dubois, D., & Zesiger, P. (2018). A cross-language study of decontextualized vocabulary comprehension in toddlerhood and kindergarten readiness. *Developmental Psychology, 54*, 1317-1333.

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, 173*, 351-370. doi:https://doi.org/10.1016/j.jecp.2018.04.012

Gambi, C., Pickering, M. J., & Rabagliati, H. (2016). Beyond Associations: Sensitivity to structure in pre-schoolers' linguistic predictions. *Cognition, 157*, 340-351.

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*, 82-90.

Hoareau, M., Yeung, H. H., & Nazzi, T. (2019). Infants' statistical word segmentation in an artificial language is linked to both parental speech input and reported production abilities. *Developmental science, 22*, e12803.

Hoff, E. (2003). The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. *Child Development, 74,* 1368-1378.

Huang, Y. T., Zheng, X., Meng, X., & Snedeker, J. (2013). Children's assignment of grammatical roles in the online processing of Mandarin passive sentences. *Journal of Memory and Language*, *69,* 589-606. doi:http://dx.doi.org/10.1016/j.jml.2013.08.002

Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., & Lyons, T. (1991). Early vocabulary growth: Relation to language input and gender. *Developmental psychology, 27*, 236-248.

Kukona, A., Fang, S.-Y., Aicher, K. A., Chen, H., & Magnuson, J. S. (2011). The time course of anticipatory constraint integration. *Cognition*, *119*, 23-42.

Leech, K. A., Rowe, M. L., & Huang, Y. T. (2017). Variations in the recruitment of syntactic knowledge contribute to SES differences in syntactic development. *Journal of Child Language*, *44*, 995-1009.

Lew-Williams, C., & Fernald, A. (2007). Young children learning Spanish make rapid use of grammatical gender in spoken word recognition. *Psychological Science, 18,* 193-198.

Lindsay, L., Gambi, C., & Rabagliati, H. (2019). Preschoolers optimize the timing of their conversational turns through flexible coordination of language comprehension and production. *Psychological Science*, *30*, 504-515.

Luke, S. G., & Christianson, K. (2016). Limits on lexical prediction during reading. *Cognitive Psychology*, *88,* 22-60.

Mahr, T., McMillan, B. T., Saffran, J. R., Weismer, S. E., & Edwards, J. (2015). Anticipatory coarticulation facilitates word recognition in toddlers. *Cognition, 142,* 345-350.

Mani, N., Daum, M. M., & Huettig, F. (2016). "Pro-active" in many ways: Developmental evidence for a dynamic pluralistic approach to prediction. *Quarterly Journal of Experimental Psychology*, *69,* 2189-2201.

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.

Marchman, V. A., & Fernald, A. (2008). Speed of word recognition and vocabulary knowledge in infancy predict cognitive and language outcomes in later childhood. *Developmental science, 11*, F9-F16.

McCauley, S. M., & Christiansen, M. H. (2019). Language learning as language use: A cross-linguistic model of child language development. *Psychological Review, 126*, 1-51.

Mirman, D. (2014). *Growth curve analysis and visualization using R*. Boca Raton, FL: CRC Press.

Nimon, K., Oswald, F., & Roberts, J. K. (2016). *Yhat: Interpreting Regression effects. R Package* Version 2.0-0. Available at: http://CRAN.R-project.org/package=yhat

Omaki, A., & Lidz, J. (2015). Linking parser development to acquisition of syntactic knowledge. *Language Acquisition, 22*, 158-192.

Peter, M. S., Durrant, S., Jessop, A., Bidgood, A., Pine, J. M., & Rowland, C. F. (2019). Does speed of processing or vocabulary size predict later language growth in toddlers?. *Cognitive Psychology, 115,* doi:10.1016/j.cogpsych.2019.101238.

Pickering, M. J., & Gambi, C. (2018). Predicting while comprehending language: A theory and review. *Psychological Bulletin, 144,* 1002-1044.

Pozzan, L., & Trueswell, J. C. (2015). Revise and resubmit: How real-time parsing limitations influence grammar acquisition. *Cognitive Psychology, 80*, 73-108.

Rabagliati, H., Gambi, C., & Pickering, M. J. (2015). Learning to predict or predicting to learn? *Language, Cognition, and Neuroscience, 31*, 94-105. doi:10.1080/23273798.2015.1077979

Ramscar, M., Dye, M., & McCauley, S. M. (2013). Error and expectation in language learning: The curious absence of mouses in adult speech. *Language, 89*, 760-793.

Ray-Mukherjee, J., Nimon, K., Mukherjee, S., Morris, D. W., Slotow, R., & Hamer, M. (2014). Using commonality analysis in multiple regressions: a tool to decompose regression effects in the face of multicollinearity. *Methods in Ecology and Evolution*, *5*, 320-328.

Reuter, T., Borovsky, A., & Lew-Wlliams, C. (2019). Predict and redicrect: Prediction errors support children's word learning. *Developmental Psychology, 55*, 1656-1665.

Roembke, T. C., & McMurray, B. (2016). Observational word learning: Beyond propose-but-verify and associative bean counting. *Journal of Memory and Language, 87*, 105-127.

Rowe, M. L. (2012). A longitudinal investigation of the role of quantity and quality of child-directed speech in vocabulary development. *Child Development*, *83*, 1762-1774.

Scottish Index of Multiple Deprivation - SIMD16 Technical Notes. (2016). Retrieved from https://www2.gov.scot/Topics/Statistics/SIMD

Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition, 128*, 302-319.

Staub, A., Grant, M., Astheimer, L., & Cohen, A. (2015). The influence of cloze probability and item constraint on cloze task response time. *Journal of Memory and Language, 82*, 1-17.

Trueswell, J. C. (2008). Using eye-movements as a developmental measure within psycholinguistics. In I. A. Sekerina, E. M. Fernandez, & H. Clahsen (Eds.), *Language processing in children* (pp. 73-96). Amsterdam: John Benjamin.

Trueswell, J. C., Sekerina, I., Hill, N. M., & Logrip, M. L. (1999). The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition*, *73*, 89-134.

Weisleder, A., & Fernald, A. (2013). Talking to children matters: Early language experience strengthens processing and builds vocabulary. *Psychological Science, 24*, 2143-2152.

Weizman, Z. O., & Snow, C. E. (2001). Lexical output as related to children's vocabulary acquisition: Effects of sophisticated exposure and support for meaning. *Developmental psychology*, *37*, 265-279.

Williams, R. H., Zimmerman, D. W., & Zumbo, B. D. (1995). Impact of measurement error on statistical power: Review of an old paradox. *The Journal of Experimental Education, 63,* 363-370.

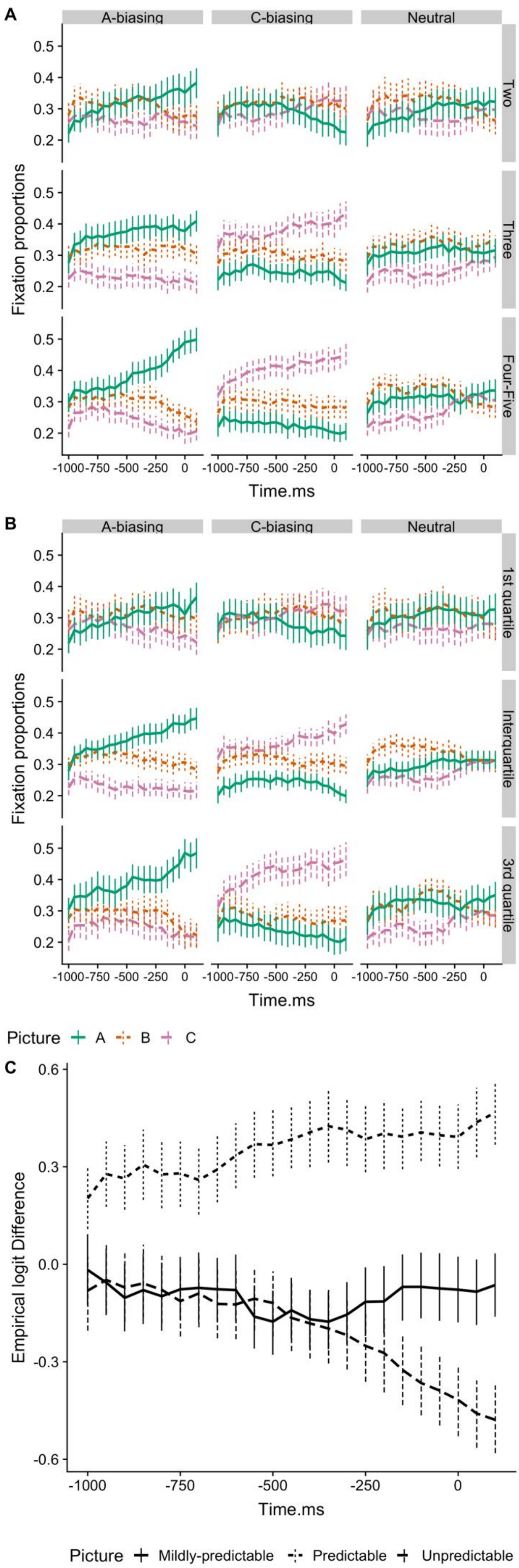
Woodard, K., Gleitman, L. R., & Trueswell, J. C. (2016). Two-and three-year-olds track a single meaning during word learning: Evidence for Propose-but-verify. *Language Learning and Development, 12,* 252-261.

List of Figures

Figure 1. Sample picture set corresponding to the sentences in Table 1. Pictures were arranged in a triangular grid as shown.

Figure 2. Gaze patterns during the prediction window. Raw fixation proportions to the three pictures as a function of context and (A) age group (two year olds, three year olds, and four-to-five year olds) or (B) quartile of the raw vocabulary measure (1^{st} quartile, interquartile range, 3^{rd} quartile). (C) Time course of the empirical log odds of looking at the predictable (fine dashed line), unpredictable (coarser dashed line), and mildly predictable

picture (solid line) while listening to predictive vs. neutral contexts. Error bars represent 95% bootstrap CI’s.


A
A-biasing
C-biasing
Neutral
Two
Three
Four-Five
Fixation proportions
Time.ms
B
1st quartile
Interquartile
3rd quartile
Picture A B C
C
Empirical logit Difference
Picture Mildly-predictable Predictable Unpredictable

1014

Figure 3 – Effect of neutral (triangles) vs. predictive (circles) contexts on the recognition of mildly-predictable pictures. Proportion of looks (time-course) as a function of age group (A) or quartiles of raw vocabulary size (B). (C) Average latency of first fixations across all children. Error bars are 95% bootstrap CIs.

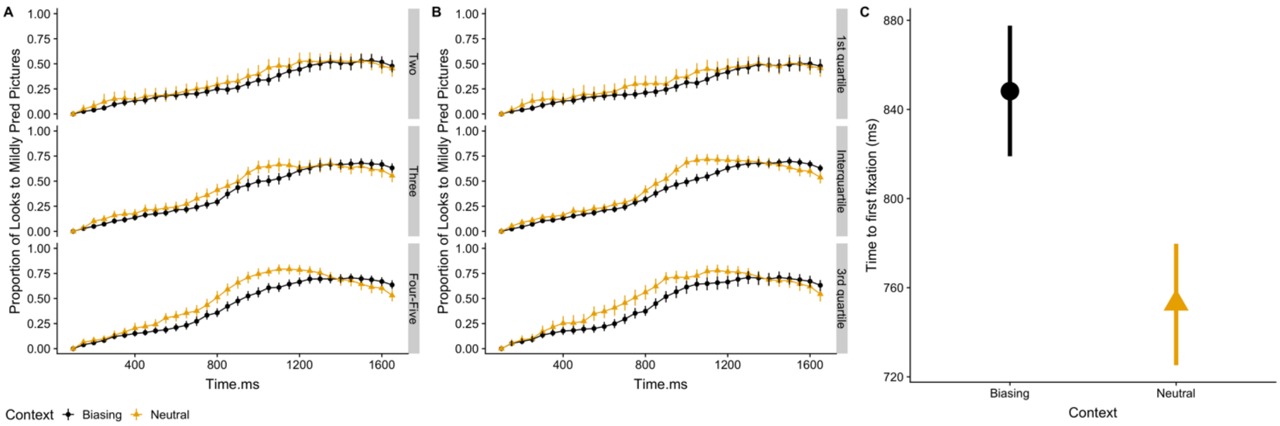


Figure 4. The cross-sectional relation between vocabulary size and: (A) the combined measure of prediction skill, (B) the predict-and-redirect measure of revision skill, (C) the time to first fixation measure of processing speed.

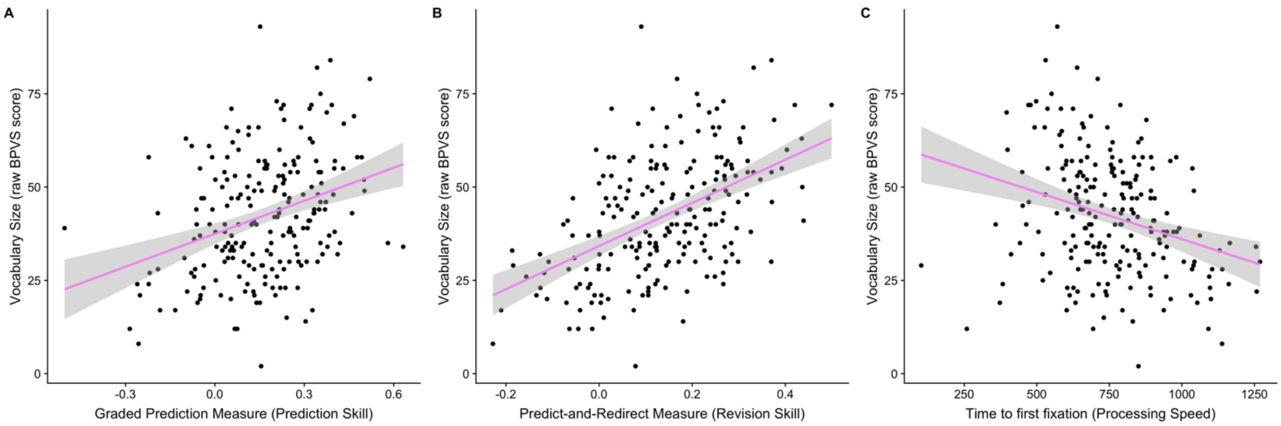
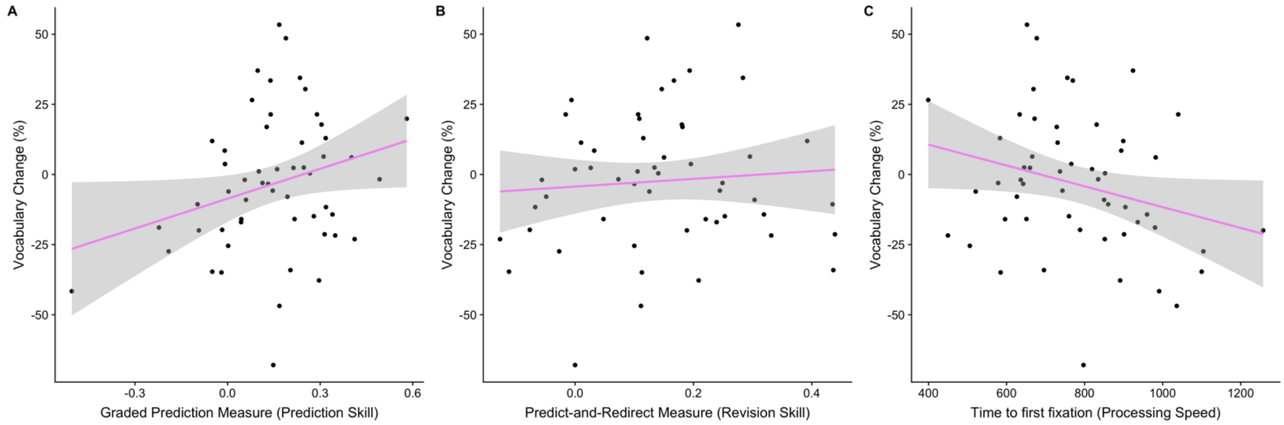


Figure 5. The longitudinal relation between the rate of vocabulary change and: (A) the combined measure of prediction skill, (B) the predict-and-redirect measure of revision skill, (C) the time to first fixation measure of processing speed.



List of Tables

Table 1. Sample sentences from an item set. Children saw a pictured bone, pair of slippers, and pair of pyjamas (as in Figure 1). See Supplementary materials, *§1* for a full item list.

Context			Final Word		
			A	B	C
Predictive	A-biasing	Alfie's dog likes to chew on the	bone	slippers	-----[a]
	C-biasing	When you go to bed, you wear	----[a]	slippers	pyjamas
Non-predictive	Neutral	Now, Craig is looking for the	bone	slippers	pyjamas

[a] Context-Final Word combinations that were not tested.

Table 2. Growth curve analysis of the prediction window. Estimate (B), standard error (SE), t value and 95% Confidence Intervals (CI) associated with key contrasts: Predictable vs. Mildly Predictable (left-hand side) and Unpredictable vs. Mildly Predictable (right-hand side). For each contrast, the model included three parameters: intercept, time, time^2. Significant parameters, i.e., those with |t|>2 (Baayen, Davidson, & Bates, 2008) are in bold.

Term	B (SE)	t	95% CI[a]
Pred – Mildly Pred	**.45(.05)**	**8.82**	**[.35,.56]**
*time	.32(.19)	1.70	[-.05,.70]
*time^2	**-.21(.11)**	**-2.01**	**[-.42,-.01]**
Unpred – Mildly Pred	**-.11(.05)**	**-2.05**	**[-.21,-.004]**
*time	**-.58(.20)**	**-2.99**	**[-.97,-.20]**
*time^2	**-.34(.10)**	**-3.24**	**[-.54,-.13]**

[a] computed with the *confint* function (method=”Wald”).

The relation between preschoolers' vocabulary development and their ability to predict and recognize words

Supplementary Materials

This document contains ancillary details about our methods as well as additional analyses. Data and scripts can be found at https://osf.io/9ckwe/.

Table of contents

1. Full list of materials and results of norming study.

Table S1. For the A-biasing (A-b) and C-biasing (C-b) conditions, we report the proportion of participants who chose the implied ordering (ABC or CBA, respectively). For the neutral condition (N), we report the highest proportion of participants that converged on the same ordering; we specify what that ordering was within brackets (e.g., BCA); in case of a tie, (---) appears instead. Proportions are based on norming study B for adults and norming study C for children (See §2 for details).

Item	Sentence	Object A	Object B	Object C	Cond	Prop. child	Prop. adult
	Alfie's dog likes to chew on the	Bone	Slippers	Pyjamas	A-b	.875	1
	When you go to bed, you wear	Bone	Slippers	Pyjamas	C-b	.750	1
	Now, Craig is looking for the	Bone	Slippers	Pyjamas	N-b	.375 (ACB)	.333 (BCA)
	After a bath, Claire wraps herself in a warm	Towel	Blanket	Pillow	A-b	.875	.833
	When you go to bed, you put your head on the	Towel	Blanket	Pillow	C-b	.875	.917
	Colin's mum will put away the	Towel	Blanket	Pillow	N-b	.500 (BCA)	.417 (BAC)
	When he wakes up, Jim opens his	Eyes	Window	Tree	A-b	.875	.750
	In the garden, grandpa likes to sit by the	Eyes	Window	Tree	C-b	.625	.750
	Tim will find the picture of the	Eyes	Window	Tree	N-b	.375 (ABC)	.583 (CBA)
	Be careful with that knife or you will cut your	Finger	Apple	Ice cream	A-b	.750	.917
	It is a hot day so Ally will eat an	Finger	Apple	Ice cream	C-b	.750	1
	Now, Bob can see the	Finger	Apple	Ice cream	N-b	.375 (BCA)	.250 (---)
	It is very cold and Lea wears her	Scarf	Glasses	Leg	A-b	.625	.917
	Sam's dad can't play football because he has broken his	Scarf	Glasses	Leg	C-b	.625	1
	Rosie is touching her	Scarf	Glasses	Leg	N-b	.375 (CBA)	.833 (CBA)
	The king's castle has a very tall	Tower	Flag	Hand	A-b	.625	.917
	Brody is saying goodbye to Mark: he's waving his	Tower	Flag	Hand	C-b	.625	.917
	Jacob will touch the	Tower	Flag	Hand	N-b	.500 (BAC)	.333 (---)
	Olivia will take a nap on the	Bed	Grass	Hair	A-b	.875	.917
	The hairdresser will cut the long	Bed	Grass	Hair	C-b	1	.917

Freddie is touching the	Bed	Grass	Hair	N-b	.750 (BAC)	.417 (BAC)
The boy is eating cereal with some	Milk	Chocolate	Letter	A-b	.750	1
James will send Santa Claus a	Milk	Chocolate	Letter	C-b	.625	.917
On the table, Sarah can see the	Milk	Chocolate	Letter	N-b	.375 (---)	.333 (ACB)
John loves racing to nursery on his	Scooter	Pony	Bunny	A-b	.625	.75
Rebecca will give a carrot to the little	Scooter	Pony	Bunny	C-b	.625	.917
Eva really likes the	Scooter	Pony	Bunny	N-b	.375 (ACB)	.417 (CBA)
At the zoo, they will see the	Elephant	Guinea Pig	Christmas tree	A-b	.750	.833
For Christmas, Mark's dad will bring home a	Elephant	Guinea Pig	Christmas tree	C-b	.750	.1
Rory is making a drawing of the	Elephant	Guinea Pig	Christmas tree	N-b	.375 (ACB)	.417 (CAB)
Amy will brush her long	Hair	Coat	Umbrella	A-b	.625	1
It might rain today: let's bring your	Hair	Coat	Umbrella	C-b	.750	1
Amy likes her mum's	Hair	Coat	Umbrella	N-b	.750 (ABC)	.667 (ABC)
The pirate will hide his treasure on the	Island	Boat	Bike	A-b	.625	1
Ryan does not like walking, he prefers to go on a	Island	Boat	Bike	C-b	.750	1
Rebecca does not like the	Island	Boat	Bike	N-b	.500 (CBA)	.417 (CBA)
Today Billie is sick, so her mum will call the	Doctors	School	Beach	A-b	.750	.833
Today, Cameron will build a sand castle at the	Doctors	School	Beach	C-b	.875	1
This morning, Charlie will go to the	Doctors	School	Beach	N-b	.375 (BAC)	.333 (---)
To make a sandwich you need two slices of bread and a slice of	Cheese	Tomato	Ball	A-b	.875	1
On the beach, Sophie will throw her sister a round	Cheese	Tomato	Ball	C-b	.625	1

Now, Isla will take the	Cheese	Tomato	Ball	N-b	.375 (CAB)	.583 (CAB)
It's getting dark and it's time to switch on the	Lamp	Oven	Window	A-b	.875	.750
It's cold and Isabella will close the	Lamp	Oven	Window	C-b	.625	.917
For the new house, Alice needs a new	Lamp	Oven	Window	N-b	.375 (CBA)	.417 (ABC)

2. Norming study methods.

We first normed the materials on adults (Norming Study A and B) and then on children (Norming Study C). Norming study A was designed to coarsely pre-screen sentence contexts for predictability using written completions, whereas Norming study B and C tested the predictability of sentence contexts in combination with the pictures that would later be used in the main experiment.

Norming Study A (Adults). We recruited 139 self-reported native speakers of English using the online platform Crowd Flower (only UK-based IP addresses were allowed). Each participant rated a minimum of 5 and a maximum of 30 randomly selected sentences, drawn from an initial pool of 60 items X 3 = 180 sentences. Sentences were accompanied by three possible completions in written form. Participants were instructed to read each sentence carefully, then order the completions from best to worst. They were encouraged to follow their first intuitions, and to “say the sentences in their head” to decide which completion sounded most natural. We discarded 18 items because either the *A-biasing* or the *C-biasing* sentence elicited the intended ordering in less than 80% of participants. Among the remaining 42 items, a large proportion of *neutral* sentences were in fact somewhat biasing towards a particular ordering. These sentences were modified in an attempt to make them more neutral, before conducting Norming study B.

Norming Study B (Adults). We recruited 36 adults using Amazon Mechanical Turk. All but 4 confirmed to be native speakers of English based in the USA (the other participants did not provide a response to these screening questions). Sentences were accompanied by pictures of possible completions. We created 3 lists, so that each participant only rated each item once, but every item was rated by 12 participants in each condition (i.e., A-biasing, C-biasing or neutral sentence). We counterbalanced the position of the objects on the screen (left-to-right ordering) between items. Six “catch” items (with obvious ordering) were included to make sure participants were paying attention. One participant gave the incorrect answer to more than 1 “catch” item (<83%) and was replaced. Six items were discarded because either the A-biasing or the C-biasing sentence elicited the intended ordering in less than 75% of participants, leaving 36 items. Again, 9 of these items did not meet the additional condition that no particular ordering should be preferred (i.e., chosen by more than 75% of participants) for the neutral sentence. These sentences were further modified, and then rated by 10 new participants recruited via Amazon Mechanical Turk; two participants were replaced because they failed to answer at least 83% of the “catch” items correctly. After modifications, only one of the neutral sentences elicited a particular ordering more than 75% of the time (see Table S1, §1).

Norming Study C (children). Finally, we collected rank-Cloze data for modified 36 items from 24 3-to-5-year-olds (M_{age} = 53 months, range [37,69], 11 males). A further 10 children were discarded for one or more of the following reasons: (1) they were bilingual with a dominant language other than English; (2) they did not follow task instructions (e.g., they always selected the pictures in the order they were presented, or deliberately selected pictures to create “silly” stories); (3) they did not complete the session.

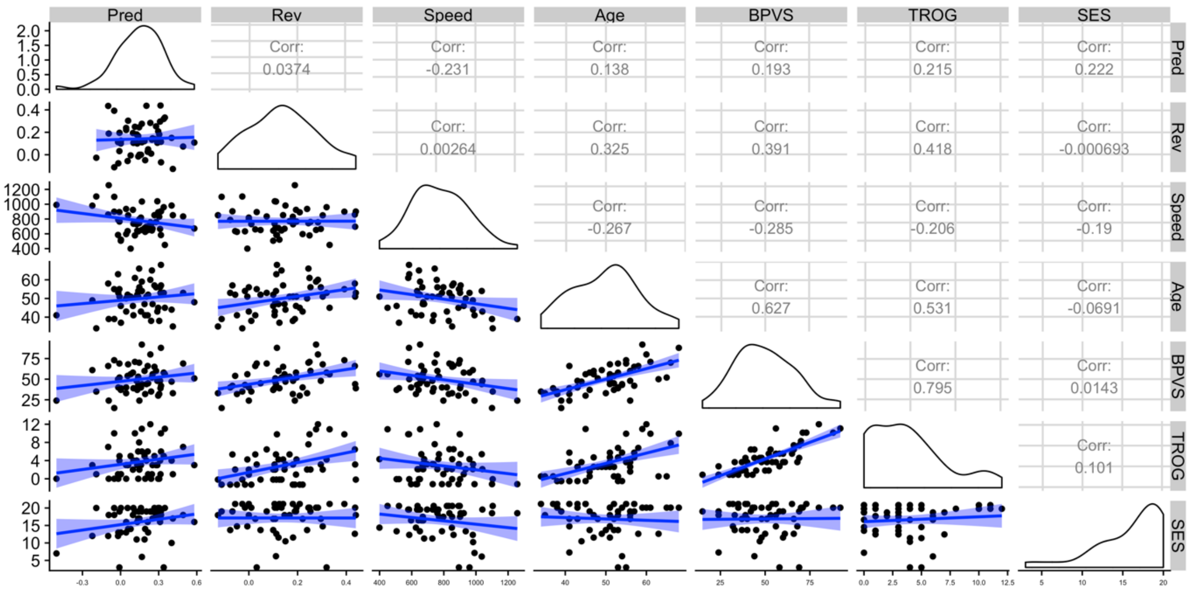
We presented the rating task as a game. The experimenter placed three boxes of different shapes and sizes in front of the child. The left-most box (labelled the “happy box”) was covered in stickers of a happy face, while the right-most box (i.e., the “sad box”) had stickers of a sad face; there were no stickers on the middle box. Children were told they would listen to stories, but these stories would all be missing the last word. The experimenter then asked for the child’s help in finding the picture that would be the best end for each story. The pictures were laid out on the table before each story, in a random order. After playing the sentence, the experimenter encouraged the child to put the best picture completion inside the “happy box”. Then she drew the child’s attention to the remaining two pictures, and after playing the story once more, asked which of the two remaining pictures would be a better completion than the other (this picture would then be put in the middle box). Given the complexity of the task, the experimenter explained it first while working through a simplified practice trial (which had an obvious implied ordering) with the child. Most children completed the practice trial correctly, but if they did not, the experimenter provided corrective feedback and explained the reasoning behind her choices using age-appropriate language.

We created 3 counterbalanced lists, so that each set of pictures was rated by 8 children in combination with each sentence, and each child only rated one set of pictures once. For each list, we used two random presentation orders (one the reverse of the other). Sentences had been pre-recorded by a female native speaker of Scottish English using natural, child-directed prosody, and were played over loudspeakers. Children were tested at the developmental lab of the Department of Psychology, University of Edinburgh, or in a quiet area at their nursery. A session lasted approximately 20 to 30 minutes. Children were allowed to take breaks at any time and were rewarded with stickers.

We selected 15 items that met the following conditions: both the *A-biasing* and the *C-biasing* sentence elicited the intended ordering at least 62.5% of the time, which is equivalent to at least 15 of the 24 children tested selecting that ordering. Two of the *non-biasing* sentences elicited a particular order more than 62.5% of the time (see Table S1), but we opted to include these items in the main experiment anyway to ensure an equal number of items per condition. In the final set of items, A-biasing sentences elicited the intended ordering (ABC) from 76% of children who took part in the norming study on average; C-biasing sentences elicited the intended ordering (CBA) from 73% of children on average; when averaged across all six possible orderings, the percentage of children who selected a given ordering for neutral sentences was 22%, while the percentage of children who converged on the most preferred ordering(s) ranged from 37.5% to 75% (average = 45%, see Table S1) for these sentences.

3. Relation between processing measures, age, vocabulary size, knowledge of grammar, and socio-economic status in the longitudinal sample.

Figure S1. Correlations between measures at Phase 2 (N = 55). Please refer to the main text for a definition of the processing measures: Pred = combined measure of graded prediction skill; Speed = measure of processing speed; Rev = measure of revision skill. The other measures are Age (months), BPVS (raw receptive vocabulary score on the British Picture Vocabulary Scale), TROG (raw grammar score on the Test for the Reception of Grammar), and SES (socio-economic status defined as the vigintile of the Scottish Index of Multiple Deprivation (2016); higher numbers indicate less deprivation).



As can be seen in Figure S1, Children's grammar knowledge was positively correlated with age (r(52)=.531, p <.001) and concurrent vocabulary size (r(52)=.795, p <.001). Interestingly, the correlations with graded prediction skill (r(52)=.215, p =.118) and processing speed (r(52)=-.206, p =.136) were in the expected direction but weak and not statistically reliable; in contrast, the correlation with revision skill was moderate and statistically significant (r(50)=.418, p<.005)[1].

However, once we controlled for age and concurrent vocabulary size in a multiple regression model, none of the processing measures explained a significant amount of variance in grammar knowledge (see Table S2 for the full model). Importantly, note that this analysis differs from the one we report in the main text for the rate of vocabulary development in the longitudinal sample (see the section *Longitudinal analysis*): since we only measured children' knowledge of grammar at Phase 2, we can only run a cross-sectional analysis for this measure. In any case, we found little evidence that variation in grammatical knowledge was explained by processing measures over and above the effects of vocabulary knowledge and age.

Table S2. Model predicting raw TROG score, as a function of the child's age in Phase 2, their concurrent raw BPVS score (centered), and the measures of graded prediction skill, revision skill, and processing speed taken at Phase 1 (transformed to *z* scores to be on a comparable scale). Significant predictors (i.e., with |t| > 2) are in bold.

Term	B (SE)	t
Intercept	3.75 (0.29)	13.04
Age	0.01 (0.05)	0.29
Vocabulary (BPVS)	**0.15 (0.02)**	**6.22**
Graded prediction skill	0.21 (0.35)	0.61

[1] We were unable to compute the revision skill measure for two participants due to missing data (see *The development of revision skills* in the main text).

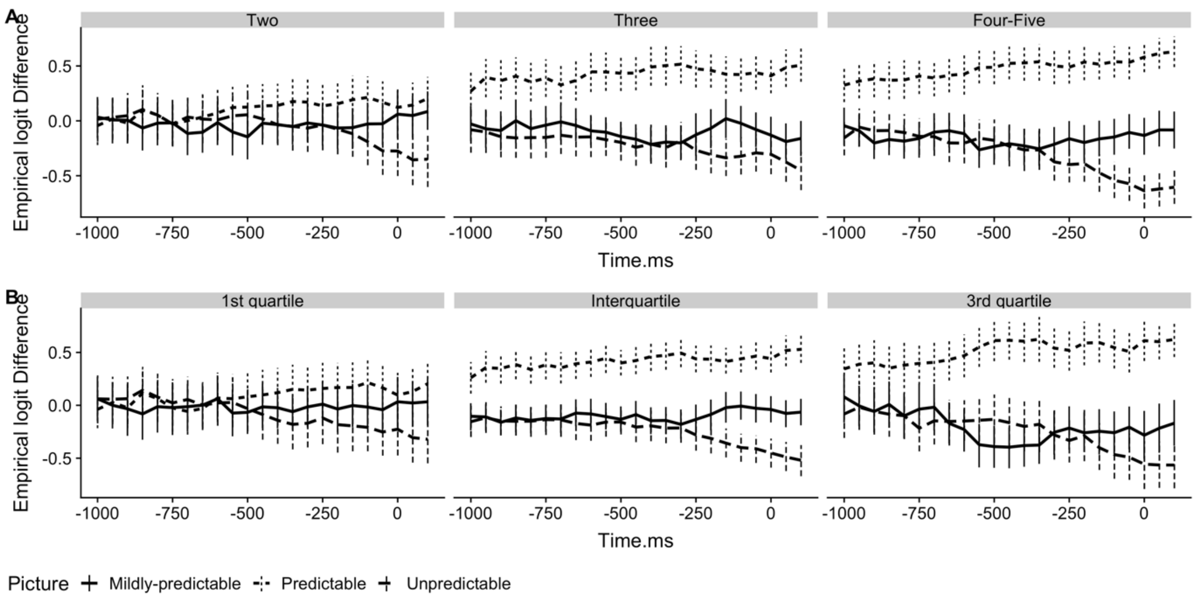
Revision skill	0.39 (0.32)	1.23
Processing Speed	0.12 (0.31)	0.40

4. Cross-sectional analyses: Graded pattern in the prediction window.

4.1. Difference curves recapitulate age and vocabulary effects observed in the raw gaze proportion data.

As noted in the main text, it is not possible to compare looks to different pictures directly (i.e., within the same condition) because this would violate the assumption of independence. Instead, we computed difference curves: after applying the *elog* transformation, we subtracted, separately for each picture, the proportion of looks to that picture after a neutral context from the proportion of looks to that picture after an A-biasing or a C-biasing context. These curves correspond to log odds of looking at that picture in one of the biasing contexts versus the neutral context. They are plotted in Figure S2 to show the same age- and vocabulary-related differences that are evident in the graphs of raw fixation proportions (Figures 2A and 2B in the main text) are also evident when we plot difference curves.

Figure S2. Difference curves (as in Figure 2C in the main text), as a function of (A) Age and (B) raw BPVS score.



4.2 By-participant growth-curve models, separately for A-biasing and C-biasing contexts.

In the main text, our growth-curve models collapsed across A-biasing and C-biasing contexts to increase the reliability of the estimates. Here, we report separate models for A-biasing and C-biasing contexts to show that (1) the results were replicated within each type of context and (2) by changing the sentential context, we could reverse children's looking preferences for the same set of pictures.

The A-biasing model compared the log odds of looking at each picture after an A-biasing context vs. a neutral context, while the C-biasing model compared the log odds of looking at each picture after a C-biasing context vs. a neutral context. Thus, we expected the A-biasing model to show that the difference curve for A pictures is higher than the difference curve for B pictures (i.e., the A–B dummy contrast should be significant), and also that the difference curve for C pictures is lower than then the difference curve for B pictures (i.e., the C–B dummy contrast should also be significant); full model in *lmer* syntax: elog (Prop. A-biasing – Prop neutral) ~ 1 + (time + time^2)*(A-B + C-B)*(Age+ Vocabulary), plus full by-participant random effects. Conversely, we expected the C-biasing model to show a higher difference curve for C pictures than B pictures, and also a lower difference curve for A than B pictures; full model: elog (Prop. C-biasing – Prop neutral) ~ 1 + (time + time^2)*(A-B + C-B)*(Age+Vocabulary), plus full by-participant random effects. Both models included age and vocabulary as (centred) covariates, so the findings we report in Table S3 below are valid for a child of average age and average vocabulary.

A-biasing model. Children were more likely to look at the highly predictable (A) than the mildly predictable (B) picture following an A-biasing context (A-B in Table S3, left panel), and this preference gradually increased over the prediction window ([A-B]**time*). Although overall they were not less likely to look at the unpredictable (C) picture than the mildly predictable (B) picture (C-B), they nevertheless became less and less likely to look at the unpredictable picture ([C-B]**time*), particularly towards the end of the prediction window, resulting in a downward-shaped curve ([C-B]*$time^2$).

C-biasing model. Children were more likely to look at the highly predictable (C) than the mildly predictable (B) picture following a C-biasing context (C-B in Table S3, right panel), and they were also less likely to look at the unpredictable (A) than the mildly predictable (B) picture (A-B). Moreover, looks to the unpredictable picture decreased over time compared to looks to the mildly predictable picture ([A-B]**time*), particularly towards the end of the time window, resulting in a downward-shaped curve ([A-B]*$time^2$). In contrast, looks to the predictable picture seemed to peak earlier and the curve had begun descending by noun onset ([C-B]*$time^2$).

Table S3. Growth-curve analysis of the prediction window, separately for A-biasing and C-biasing contexts. Estimates (B), standard errors (SE), t values and 95% Confidence Intervals (CI) associated with key contrasts in the A-Biasing model (left) and the C-biasing model (right); the contrasts are: A vs. B pictures (A-B) and C vs. B pictures (C-B). For each contrast, the model includes three parameters, for the intercept, first order time term (*time) and second order time term (*time^2). See main text for the interpretation of the different parameters. Significant parameters (|t|>2) are highlighted in bold.

		A-biasing model			C-biasing model		
Term		B (SE)	t	95% CI[a]	B (SE)	t	95% CI[a]
A – B		**.33(.07)**	**4.98**	**[.20,.45]**	**-.18(.07)**	**-2.65**	**[-.31,-.05]**
	*time	**.58(.25)**	**2.29**	**[.08,1.07]**	**-.58(.26)**	**-2.20**	**[-1.10,-0.06]**
	*time^2	-01(.15)	-0.08	[-.30,.28]	**-.33(.16)**	**-2.04**	**[-.64,-.01]**
C - B		-.03(.06)	-0.50	[-.16,.10]	**.58(.07)**	**8.44**	**[.45,.72]**
	*time	**-.59(.25)**	**-2.30**	**[-1.08,-0.09]**	.07(.24)	0.30	[-.41,.55]
	*time^2	**-.35(.15)**	**-2.32**	**[-.65,-.06]**	**-.41(.16)**	**-2.56**	**[-.73,.09]**

[a] computed with the *confint* function (method=”Wald”).

4.3 By-item growth-curve models (collapsing across A-biasing and C-biasing contexts).

The models reported in this section have the same form as the ones reported in the main text (i.e., they collapse across A-biasing and C-biasing contexts), but the data were averaged over participants to obtain by-item estimates (rather than vice versa). Since age and vocabulary are participant-specific measures, they were not entered into by-items models. Table S4 shows that by-item analyses largely confirmed by-participant analyses, though the effects were generally weaker and only reliable on selected terms (highlighted in bold in the table). Importantly, however, there was evidence for both an overall preference for predictable over mildly predictable pictures (Pred - Mildly Pred) and a gradual decrease in looks to the unpredictable (compared to the mildly predictable) picture over time ([Unpred – Mildly Pred] * *time*).

Table S4. Growth-curve analysis of the prediction window, with items as the source of random variation. This table corresponds to Table 2 in the main text, except that it shows analyses over items, rather than over participants.

Term	B (SE)	t	95% CI[a]
Pred – Mildly Pred	**.53(.08)**	**6.59**	**[.37,.68]**
*time	.38(.24)	1.55	[-.10,.85]
*time^2	-.21(.17)	-1.24	[-.54,.12]
Unpred – Mildly Pred	-.12(.07)	-1.70	[-.25,.02]
*time	**-.69(.29)**	**-2.40**	**[-1.26,-.12]**
*time^2	-.35(.20)	-1.70	[-.75,.05]

[a] computed with the *confint* function (method=”Wald”).

4.4 Interactions with age/vocabulary in the by-participant growth-curve models, collapsing across A-biasing and C-biasing contexts.

In the main text, we did not discuss the interactions between the covariates age and vocabulary and the other parameters of the growth-curve model modelling looks during the prediction window. These interactions are reported in Table S5 and discussed below.

Table S5. This table complements Table 2 in the main text, reporting interactions between the parameters shown in Table 2 and either concurrent Age (in months; left) or Vocabulary (raw BPVS score; right), both centered. Significant interactions are highlighted in bold.

	Interactions with Age			Interactions with Vocabulary		
Term	B (SE)	t	95% CI[a]	B (SE)	t	95% CI[a]
Pred – Mildly Pred	.01(.01)	1.22	[-.01,.03]	.01(.01)	1.43	[-.003,.02]
*time	-.03(.03)	-0.92	[-.09,.03]	.03(.02)	1.62	[-.01,.07]
*time^2	.03(.02)	1.56	[-.01,.06]	-.02(.01)	-1.94	[-.04,.003]
Unpred – Mildly Pred	-.01(.01)	-1.41	[-.03,0.005]	.01(.01)	1.04	[-.005,.02]
*time	-.05(.03)	-1.64	[-.11,.01]	.03(.02)	1.70	[-.01,.07]

*$time^2$	**.04(.02)**	**2.46**	**[.01,.08]**	**-.03(.01)**	**-2.95**	**[-.05,-.01]**

[a] computed with the *confint* function (method="Wald").

Perhaps surprisingly, there was no indication that parameters' estimates varied with either age or vocabulary, with the exception of the parameter capturing the decrease in looks to unpredictable pictures towards the end of the prediction window (in Table S5: [Unpred – Mildly Pred] *$time^2$). The model indicated that this decrease tended to be steeper (more negative) in children with larger vocabulary, but shallower (more positive) in older children. In contrast, neither age nor vocabulary affected the magnitude or time-course of the preference for highly predictable over mildly-predictable pictures (see the top three rows of Table S5). Note that the models' findings are not fully reflected in Figure S2 because the model captures the effect of age while controlling for vocabulary, and vice versa, whereas the figure shows the effect of age ignoring variability in vocabulary size, and vice versa.

These initial findings may suggest that the ability to differentiate mildly predictable from unpredictable pictures is associated with more advanced linguistic skills (over-and-above age differences) in our cross-sectional sample. Accordingly, when we compared the fit of the full model (including interactions with both age and vocabulary) to the fit of the model including only interactions with age (using a log-likelihood ratio test as implemented by the function *anova()* in R, package *lme4*), we found that adding vocabulary to the model improved fit somewhat ($\chi2(9) = 17.46$, $p = .042$). Further, we found that the increase in fit was due to interactions between vocabulary and the dispreference for unpredictable pictures ($\chi2(3) = 10.49$, $p = .02$), whereas including interactions between vocabulary and the preference for predictable pictures did not add to the fit of the model ($\chi2(3) = 5.14$, $p = .162$).

However, these findings should be treated with caution, for three reasons. First, vocabulary was (unsurprisingly[2]) strongly correlated with age ($r(213) = .803$, $p<.001$), but the relation between age and raw vocabulary size in our sample could be more complex than a simple linear relation, and this might help explain why age and vocabulary seemed to be related to the dispreference for unpredictable pictures in opposite ways. Second, when we re-fit the model to include either only interactions with age or only interactions with vocabulary (i.e., elog(Prop. Predictive) – elog(Prop. neutral) ~ 1 + (time+time2)*(Predictable-Mildly predictable + Unpredictable-Mildly predictable)*[Age or Vocabulary], plus maximal by-participant random effects), we confirmed what is evident in Figures 2A and S2A and 2B and S2B, i.e. that children's prediction skills improve with both age and vocabulary, respectively. More specifically, we found that children's preference for predictable pictures grew significantly stronger with age (intercept: $t= 3.96$, other interactions $|t|< 1$) and vocabulary size (intercept: $t = 4.04$, other interactions $|t|<1.50$). In contrast, however, we did not find statistically significant evidence for age or vocabulary-related differences in children's ability to distinguish between unpredictable and mildly predictable pictures (all $|t|$'s < 1.7). Third, when we correlated vocabulary size with measures of prediction skill based on raw data from the last 400ms of the prediction window (see §4.5 below), we found no evidence for a relation between the dispreference for unpredictable pictures and vocabulary size. This suggests that the relation between vocabulary size and the [Unpred – Mildly Pred] *$time^2$ parameter in the model (see Table S5) may reflect individual differences in the shape of the curve representing the decrease in looks to unpredictable pictures towards the end of the prediction window, rather than differences in the ability to distinguish between mildly predictable and unpredictable pictures *per se*.

In sum, while the major locus of measurable individual differences was in increased differentiation of the two most predictable continuations, once age-related effects were accounted for, more advanced linguistic abilities seemed to be most associated with the time-course with which

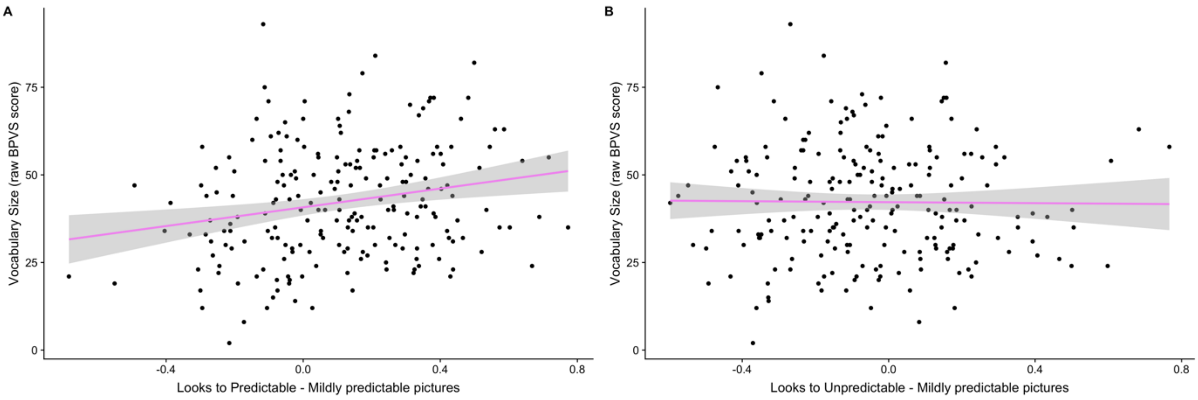
[2] The strong correlation between age and vocabulary size is unsurprising given we used raw vocabulary scores, but recall standardized BPVS scores were not available for children below the age of three.

children directed their attention away from unpredictable pictures, but the functional significance of this latter finding is unclear.

4.5 Relation between vocabulary size and the (raw) preference for predictable pictures / (raw) dispreference for unpredictable pictures.

Figure S3 below should be compared to Figure 4A in the main text, which shows the cross-sectional relation between vocabulary size at Phase 1 and the combined measure of graded prediction skill. While that relation was found to be positive and significant, the relation between vocabulary size and the degree to which children preferred to look at pictures that were highly predictable given the context over those that were only mildly predictable was significantly positive, but weaker (r(213) = .214, p <.005; see Figure S3, panel A). Moreover, the relation between vocabulary size and the dispreference for unpredictable pictures compared to mildly predictable pictures was not significant (r(213) = -.011, p >.250). Similarly, the preference measure was related to age at Phase 1 (r(213) = .193, p <.005), though not as strongly as the combined measure (see main text), while the disprefrence measure was not (r(213) = -.064, p >.250).

Figure S3. The cross-sectional relation between vocabulary size in Phase 1 (raw BPVS score) and (A) the raw preferences for predictable vs. mildly-predictable pictures and (B) the raw dispreference for unpredictable vs. mildly predictable pictures.



5. Cross-sectional analyses: The cost associated with disconfirmed predictions - interactions with age and vocabulary.

We explored how the hindering effect of inaccurate predictions changed with age and vocabulary. The full model including both age and vocabulary (see Table S6) revealed no significant age or vocabulary-related differences to the hindering effect of disconfirmed predictions. Moreover, vocabulary did not explain any additional variance over-and-above the effect of age, as adding vocabulary to a model that only included age did not significantly improve fit ($\chi^2(2) = 3.25$, p = .197). However, when we fit separate models including only age (Table S7) or only vocabulary (Table S8), we found that the effect of disconfirmed predictions grew stronger with increasing age (t = -2.62) and vocabulary (t = -2.82), confirming the visual trends in Figure 3 (3A and 3B, respectively) in the main text. So, although it is unclear what drives these individual differences (i.e., vocabulary or other skills that change with age), it is clear that the hindering effect of disconfirmed predictions increases during the preschool years.

Table S6. Model summary capturing the cost associated with a disconfirmed prediction. The effect of Context compares the time to first fixation to a mildly predictable picture after a neutral context and after a context predictive of a different picture; this model includes Age and Vocabulary as (centered) covariates. Significant predictors are highlighted in bold. Model formula: Latency ~ 1 + Context *(Age

+ Vocabulary), plus maximal random effects by item, and random intercepts by participants (by-participant slopes for Context were estimated to be close to zero and dropped for convergence)

Term	B (SE)	t	95% CI[a]
Context	**-95.51 (25.28)**	**-3.78**	**[-145.06,-45.96]**
Age	-1.07(1.70)	-0.63	[-4.40,2.25]
Vocabulary	-1.49(1.09)	-1.36	[-3.63,0.65]
Context * Age	-2.61(3.29)	-0.79	[-9.06,3.84]
Context * Vocabulary	-2.52(2.12)	-1.19	[-6.67,1.63]

[a] computed with the *confint* function (method="Wald").

Table S7. Model summary capturing the cost associated with a disconfirmed prediction. This model includes only Age as a (centered) covariate. Model formula: Latency ~ 1 + Context *Age, plus maximal random effects by item, and random intercepts by participants.

Term	B (SE)	t	95% CI[a]
Context	**-95.38 (25.40)**	**-3.76**	**[-145.16,-45.60]**
Age	**-2.81(1.08)**	**-2.59**	**[-4.93,-0.68]**
Context * Age	**-5.53(2.11)**	**-2.62**	**[-9.66,-1.40]**

[a] computed with the *confint* function (method="Wald").

Table S8. Model summary capturing the cost associated with a disconfirmed prediction. This model includes only Vocabulary (BPVS score) as a (centered) covariate. Model formula: Latency ~ 1 + Context Vocabulary, plus maximal random effects by item, and random intercepts by participants.

Term	B (SE)	t	95% CI[a]
Context	**-95.55 (25.23)**	**-3.79**	**[-144.99,-46.11]**
Vocabulary	**-2.02(0.70)**	**-2.89**	**[-3.40,-0.65]**
Context *Vocabulary	**-3.82(1.36)**	**-2.82**	**[-6.47,-1.16]**

[a] computed with the *confint* function (method="Wald").

6. Comparison between the distributions of vocabulary (Figure S4) and age (Figure S5) in the cross-sectional sample and the longitudinal subsample

Figure S4. Distribution of vocabulary scores (raw BPVS score) at Phase 1 for children tested in Phase 1 only (orange bars) and those that were later retested in Phase 2 (subsample, yellow bars).

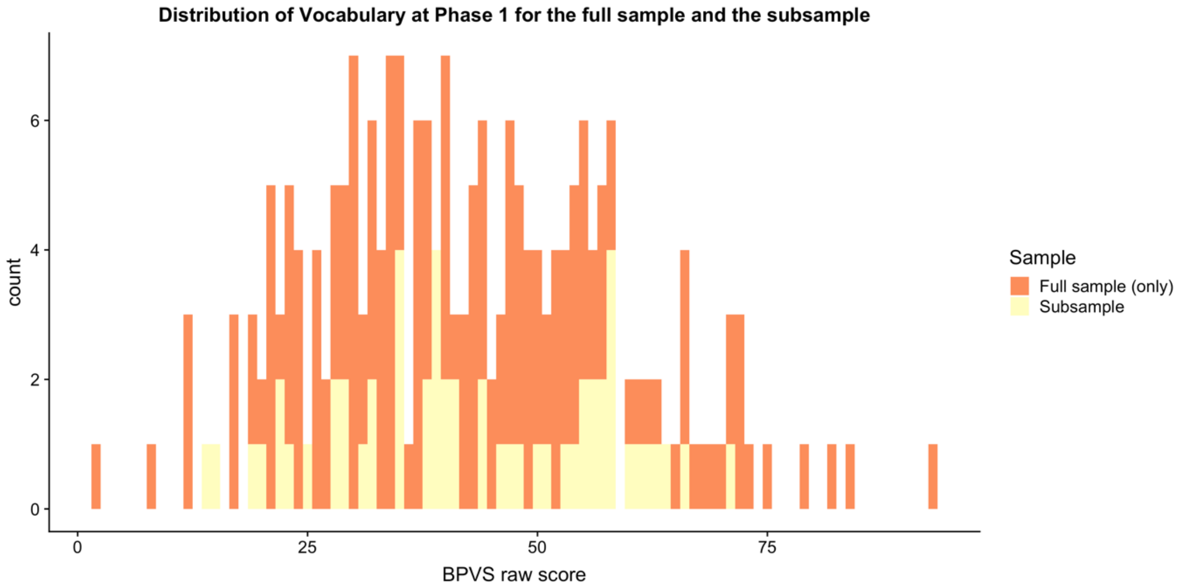
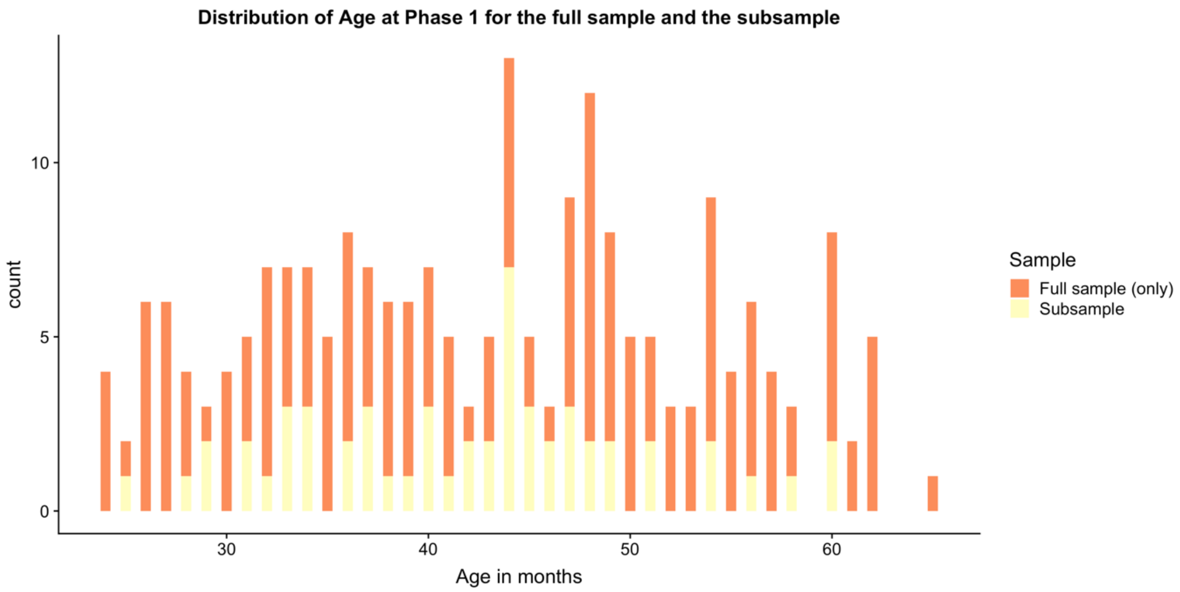


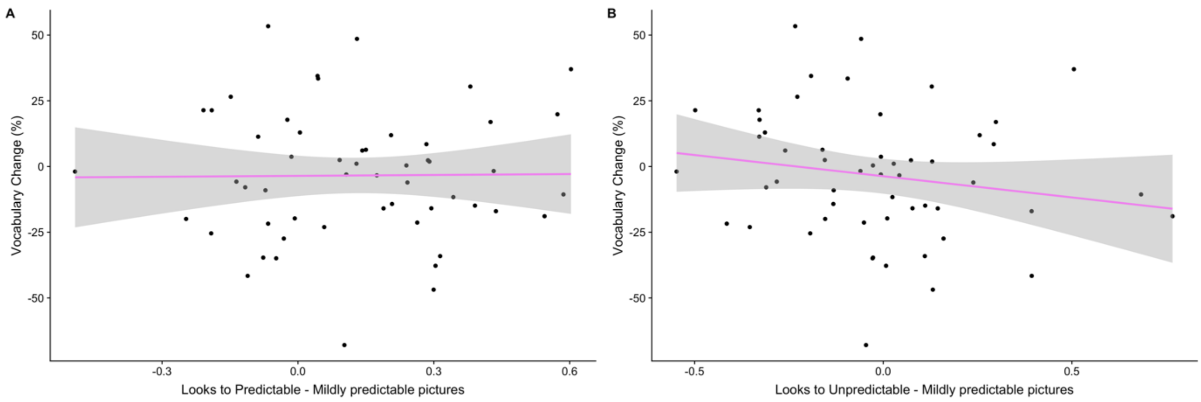
Figure S5. Distribution of age (in months) at Phase 1 for children tested in Phase 1 only (orange bars) and those that were later retested in Phase 2 (subsample, yellow bars).



7. Longitudinal analyses: Relation between vocabulary development and prediction skills.

The combined measure of graded prediction skill was a significant predictor of inter-individual variability in the rate of vocabulary development (see *Longitudinal analysis* in the main text). In contrast, the component measures (i.e., the preference for predictable and the dispreference for unpredictable pictures) were not. The preference for predictable over mildly-predictable pictures (computed over the last 400ms of the prediction window) did not predict the rate of vocabulary development when entered in a linear regression model (as in the analyses reported in the main text, we scaled the preference measure before entering it into the model, and we controlled for vocabulary size at Phase 1, centered): B = .61, SE = 3.45, t =.18. Similarly, the dispreference for unpredictable compared to mildly predictable pictures, computed over the same time window, also did not explain any variance in the rate of vocabulary development (analysis as above): B = -4.16, SE -3.39, t = -1.23. See Figure S6.

Figure S6. The relation between the rate of vocabulary change (%) and (A) the preference for predictable over mildly-predictable pictures in the last 400ms of the prediction window, (B) the dispreference for unpredictable relative to mildly-predictable pictures in the last 400ms of the prediction window.



8. Longitudinal analyses: Relation between prediction skill, revision skill and processing speed and the rate of vocabulary change (%), while controlling for Age in Phase 1

The longitudinal analyses reported in the main text controlled for vocabulary size (raw BPVS score) in Phase 1. Below, we report similar analyses but using age at Phase 1 as the control variable.

When controlling for age instead of vocabulary at Phase 1, the measure of revision skill remained unrelated to the rate of vocabulary change ($p>.250$). In contrast, both processing speed (B = -6.13, SE = 3.42, t=-1.79, p = .079) and the combined measure of graded prediction skill (B = 6.32, SE = 3.32, t = 1.905, p = .062) were marginally related to the rate of vocabulary change. Importantly, although in a multiple regression model including both measures, neither prediction (B = 5.33, SE = 3.35, t = 1.59, p = .118) nor processing speed (B = -5.03, SE = 3.44, t = -1.46, p =.151) were significant predictors of the rate of vocabulary change, the combined measure of fluent language processing improved model fit significantly compared to a baseline model including only age at Phase 1 ($F(1, 51) = 5.95$, p = .018), and the model including it explained a significant amount of variation in vocabulary development (R^2 = .119, $F(2,51) = 3.43$, p = .04).

9. Longitudinal analyses: Chronological age and linguistic age (expressed as a percentage increment of chronological age) for each child.

Table S9. Chronological age (Age) and Linguistic Age (expressed as a percentage increment of chronological age) for each child in the longitudinal subsample (N = 54) at each testing point (Phase 1 and Phase 2); Vocabulary Change (Voc Change, %) is obtained by subtracting Linguistic Age Phase 1 from Linguistic Age Phase 2.

Age Phase 1	Age Phase 2	Linguistic Age Phase 1 (as a % of Age Phase 1)	Linguistic Age Phase 2 (as a % of Age Phase 2)	Voc Change (%)
43	52	-13.95	-17.31	-3.36
42	51	-11.90	21.57	33.47
46	56	30.43	-37.50	-67.93
39	48	5.13	25.00	19.87
45	54	-28.80	-7.41	21.39
43	53	4.65	-15.09	-19.74
45	55	44.44	56.36	11.92
44	53	4.35	-16.98	-21.33
41	49	65.85	46.94	-18.91
37	46	18.91	-6.52	-25.43
37	44	18.92	25.00	6.08
54	61	40.74	37.70	-3.04
54	60	22.22	56.67	34.45
51	58	45.10	34.48	-10.62
38	45	68.42	46.67	-21.75
36	43	2.77	4.65	1.88
42	51	-7.14	9.80	16.94
40	50	-5.00	-22.00	-17.00
40	47	2.50	51.06	48.56
56	63	28.57	-6.35	-34.92
46	54	41.30	-5.56	-46.86
44	51	34.09	0.00	-34.09
44	51	15.90	19.61	3.71
40	47	10.00	40.43	30.43
44	52	25.00	23.08	-1.92
48	56	41.67	26.79	-14.88
44	53	54.54	52.83	-1.71
37	46	10.81	47.83	37.02
48	57	-4.16	-15.79	-11.63

47	56	-6.38	0.00	6.38
34	43	20.59	4.65	-15.94
34	43	8.82	30.23	21.41
29	38	-3.45	7.89	11.34
32	41	6.25	7.32	1.07
28	37	25.00	18.92	-6.08
49	55	38.78	47.27	8.49
29	35	34.48	11.43	-23.05
33	38	-15.15	2.63	17.78
34	41	29.41	-12.20	-41.61
33	41	3.03	-4.88	-7.91
44	52	15.90	0.00	-15.90
33	41	87.88	90.24	2.36
47	53	42.55	28.30	-14.25
44	51	65.90	56.86	-9.04
45	51	57.78	84.31	26.53
36	45	22.22	-15.55	-37.77
49	57	34.69	35.09	0.40
60	68	40.00	52.94	12.94
31	39	9.68	-10.26	-19.94
31	39	6.45	-28.21	-34.66
25	34	48.00	20.59	-27.41
58	65	-8.62	-6.15	2.47
60	66	30.00	24.24	-5.76
51	59	31.37	84.75	53.38