Dynamics of context-driven lexical activation in children and adults

Eun-Kyoung Rosa Lee[a,*], Katherine Howitt[a], London Dixon[a], Tal Ness[a], Masato Nakamura[a], Colin Phillips[a]

[a]University of Maryland, College Park, MD 20742, USA

*Corresponding author (ekleesla@umd.edu)

Abstract

This study investigates whether children share the same predictive mechanism as adults, where multiple candidates for upcoming input are activated in parallel based on context. A total of 60 school-aged children (ages 4-12, mean = 9) and 152 adults performed a speeded cloze task in a live museum setting, as part of their experience at a language-focused museum. Two key patterns emerged in children and adults: high-cloze responses were produced more quickly than low-cloze responses, and highly constraining contexts yielded faster responses than less constraining contexts. These findings support a shared race-like process in generating next-word predictions, where multiple candidates are activated based on context and compete toward an activation threshold (Staub et al., 2015). Despite the shared mechanism, children and adults show some differences in speeded cloze response patterns, which may reflect that children generate different response candidates. Analyses of children's data motivate the use of mean response times, rather than modal cloze probability, as a measure of sentence constraint. Implications for how cloze probability and cloze response latencies map onto the underlying cognitive processes involved in prediction are discussed.

1. Introduction

A large body of work in psycholinguistics has shown that adult comprehenders actively anticipate upcoming input during real-time sentence comprehension (Altmann & Mirković, 2009; Van Petten & Luka, 2012; Dell & Chang, 2014; Huettig, 2015; Kuperberg & Jaeger, 2016; Pickering & Garrod, 2013; Pickering & Gambi, 2018; Federmeier, 2007, 2022). Supporting evidence comes from various measures, including early looks toward targets in visual world eye-tracking (Altmann & Kamide, 1999; Kamide et al., 2003), facilitated reading times in eye-tracking while reading or self-paced reading paradigms (Ehrlich & Rayner, 1981; for a review see Staub, 2015), and brain responses that are sensitive to the predictability of the target input (e.g., Kutas & Federmeier, 2000; DeLong et al., 2005). While the exact mechanisms underlying how expectations are formed are debated, it has been argued that adult comprehenders generate multiple possible candidates for upcoming words in parallel (Roland et al., 2012; Staub et al., 2015; Frisson et al., 2017; Ness & Meltzer-Asscher, 2021), instead of limiting predictions to a single most highly probable candidate. The parallel activation of multiple candidates has been a mechanism proposed to explain key empirical findings, including systematic patterns in how quickly people produce responses in a speeded cloze paradigm (Staub et al., 2015).

Studies have shown that children exhibit anticipatory behavior similar to adults during real-time sentence comprehension (Borovsky et al., 2012; Nation et al., 2003; Mani & Huettig, 2012), but whether the underlying predictive mechanism also involves activating multiple lexical candidates in parallel is less clear. The majority of evidence supporting children's ability to form expectations for upcoming input comes from the visual world paradigm, where increased looks to a target picture before it is explicitly mentioned in the linguistic input is taken to reflect anticipatory behavior. Despite its many benefits, the method has limitations in probing the range of hypotheses children consider during prediction, or whether they consider multiple candidates simultaneously, and if so, whether they generate similar sets of candidates to adults. In the present study, we examined whether children share the same underlying predictive mechanism as adults, specifically whether they exhibit adult-like patterns in a speeded cloze paradigm which reflect the dynamics of competitive activation of multiple candidates.

1.1 Underlying mechanisms in prediction: Sampling from a probability distribution, or competition among multiple candidates?

The term *prediction* has been used in the psycholinguistics literature to refer to different processes with different assumptions about the underlying mechanisms. The present study views prediction in the form of context-driven pre-activation (DeLong et al., 2005; Kutas et al., 2011; Staub et al., 2015; Federmeier, 2022). Adult comprehenders can use contextual information in combination with their prior knowledge of the world and linguistic constraints to probe long-term memory and activate relevant items that will help them to process upcoming linguistic input. This kind of pre-activation can occur at multiple levels, including semantic features (Federmeier & Kutas, 1999) and lower-level phonological forms (Mani & Plunkett, 2010; Gambi et al., 2018). Here, we focus on the activation of lexical candidates that are considered to be a good 'fit' to the preceding context. Researchers have often measured this goodness-of-fit by using the cloze probability of a word given a context, where cloze probability is calculated by how frequently a word is produced as a continuation of a given sentence fragment within a group of people. Higher cloze probability is taken to indicate a better fit of the word to the context, relative to other potential candidates that have a lower cloze probability.

While cloze probability has been widely used to represent how predictable a particular word is based on its preceding context, exactly how it maps onto the actual strength of a candidate in relation to other potential candidates is less clear. Consider two possibilities that have been proposed in the literature, regarding the connection between cloze probability and the underlying mechanism that leads to responses with different cloze probabilities for a given context. One view is that cloze probability is derived from a process where each individual samples from a subjective probability distribution (Smith & Levy, 2011). Given world and linguistic constraints, people have prior distributions of likely continuations for a context, and in a cloze task, they simply sample from this subjective probability distribution and produce a cloze response. In this way, the proportion of responses produced for a given context directly reflects the probability distribution of those candidates. Although cloze probabilities can diverge from actual corpus-based conditional probabilities, potentially due to biases in people's subjective prior distributions (Smith & Levy, 2011), the key claim of this approach is that cloze probability is a direct reflection of how good a candidate is given a context, because people simply sample from a subjective probability distribution when they produce responses in a cloze task.

A different view on cloze probability is offered by Staub et al. (2015), where it is claimed that cloze probability is not derived by sampling from a probability distribution, but rather is an output of a competitive race process: multiple candidates gain activation based on context and the candidate to reach an activation threshold first gets produced. Unlike Smith & Levy's sampling (2011) model, Staub et al.'s (2015) race model does not view cloze probability as a direct read-out of how likely a word is given a context. Different candidates accumulate activation and race toward the threshold with different speeds, and the primary relationship is between a candidate's strength and the speed of activation. Under this theory, the time it takes for a response to get produced, or the cloze reaction time (RT), is a more direct reflection of goodness-of-fit than cloze probability, although it is not a direct reflection, since RTs are only observed for winners, i.e., responses that are produced. The key point is that the speeds of candidates are fed into the race process, and cloze probabilities reflect win proportions. In Staub et al.'s (2015) race simulations, the strength of each candidate in a given context is operationalized through a mean finishing time distribution, where this distribution plus some random noise together yield slightly different outcomes on multiple trials of the same race. A candidate's own speed and the speed of its competitors jointly determine the probability of how often the candidate wins the race, represented by cloze probability[1].

An important distinction between the sampling account and the race model is that the race model offers a direct explanation for systematic timing patterns found in adults' speeded cloze responses which are not straightforwardly explained by the sampling account. In Staub et al.'s (2015) speeded cloze paradigm, participants read sentence fragments presented in rapid serial visual presentation (RSVP) and produced cloze responses under time pressure. The authors discovered two key patterns in speeded cloze response latencies. The first was that high-cloze responses were produced, on average, more quickly than low-cloze responses. The race model naturally predicts this relationship between cloze probability and response times, where faster candidates have a better chance of winning the race and getting produced than candidates that are slow. Therefore, more frequently winning candidates should have faster finishing times. The

[1] Staub et al. (2015) make no assumptions about direct influence among competitors on their activation profiles, e.g., via lateral inhibition. In this model, the effect of different competitors on each other is limited to determining whether or not they are winners. We follow this assumption here, but see Nakamura (2023) for evidence for lateral inhibition effects in a speeded cloze paradigms.

sampling account proposed by Smith & Levy (2011) does not make any predictions regarding response latencies, as the production of a cloze response is a result of a one-time sampling from a probability distribution. However, since the original claim was proposed to explain offline cloze probabilities, the account could be extended to explain speeded cloze response latency patterns. The relationship between the subjective probability distribution and cloze probability could be used to predict a similar relationship between subjective probability and response latencies. If there is a core notion of strength of fit between a context and a candidate, this should directly impact both the probability of sampling and the speed of the response. In this way, the account would predict high-cloze responses to be faster than low-cloze responses.

The second key finding in adults' speeded cloze response latencies observed by Staub et al. (2015), however, more clearly favors the race model over the sampling account. The authors found that responses are generally faster when produced in a highly constraining context than in a low constraining context, where context constraint was measured by the cloze probability of the modal response, i.e., the most frequent response given a context. This pattern naturally follows from a competitive race process, where candidates that compete with faster competitors can only beat the strong competitors when they are themselves relatively fast. In other words, a candidate is only produced when it is activated faster than its competitors, and therefore, a candidate with faster competitors needs to be faster to win than a candidate with slower competitors. The relationship between contextual constraint and speeded cloze response latencies is not predicted by the sampling account, because a candidate's strength (and hence, presumably, its speed of generation) is taken to be directly reflected in its subjective probability. There is no parallel activation of multiple candidates, and hence no reason for the speed of a response to be affected by the strength of alternatives. The fact that responses are produced more quickly when their alternatives are stronger does not follow without additional assumptions from the sampling model.

Based on these empirical findings, there is strong support for the claim that adult comprehenders actively consider multiple possible candidates that fit the preceding context in parallel, and that there is competition among those generated candidates which affects how likely a candidate gets produced as a cloze response. In the next section, we discuss the possibility of whether children also utilize a race mechanism in predicting upcoming linguistic input during real-time sentence processing.

1.2 Children's predictive processing

Research on children's predictive processing has mostly focused on whether or not children predict upcoming input like adults. Many studies report findings indicating that children from a very young age show predictive behaviors similar to adults during real-time comprehension. One common method used to study prediction in children is the visual world eye-tracking paradigm, where the proportion of looks toward different objects displayed on a visual scene is measured while a sentence is aurally presented. Early increased looks toward a target object before it is mentioned in the sentence is taken as evidence for prediction, or anticipation of the target before it is explicitly provided in the linguistic input. Previous studies using the visual world paradigm have found that children as young as two years old exhibit adult-like anticipatory looking patterns (Nation et al., 2003; Mani & Huettig, 2012; Borovsky et al., 2012). For example, Borovsky and colleagues (2012) aurally presented sentences like, *The pirate chases the ship*, while 4-to-5-year-old children viewed a display of four objects: a treasure box, a pirate ship, a cat, and a bone. Upon hearing *chases*, the proportion of looks toward the correct object, the pirate ship, increased in comparison to the other objects. This was taken as evidence that young children are capable of using their world knowledge about pirates and a chasing event to quickly start looking toward the object representing a likely patient given the context. Similar results were found with 2-year-olds (Mani & Huettig, 2012) and 10-to-11-year-olds (Nation et al., 2003). Studies have also shown that children actively use syntactic structure (Gambi et al., 2016) and morpho-syntactic cues that are available in their language for prediction, such as grammatical gender or number agreement (Lew-Williams & Fernald, 2007; Lukyanenko & Fisher, 2016; Özge et al., 2019, 2022; Aumeistere et al., 2022; Brown et al., 2022; Smolík & Bláhová, 2022).

The previous findings highlight parallels between children's and adults' predictive processing primarily based on measures like eye-movements and responses to visually presented stimuli. The visual world paradigm, as described above, is a popular experimental method for probing children's predictive behavior due to its many advantages. It allows the experimenter to control the kind of context that is available to the participant. It is easily combined with auditory sentence presentation, and no reading is required. It provides moment-by-moment information about when the context starts to affect eye-movements. It also makes it possible to observe effects that occur before the target is present in the linguistic input, indicating anticipatory behavior. Predictive eye-movements can therefore tell us how efficiently children can analyze the given linguistic context and combine their analysis with prior world knowledge and awareness of the current visual scene to generate expectations for upcoming input.

Despite these benefits, the visual world paradigm, along with other comprehension measures like EEG/ERPs, has limitations in probing whether there is parallel, competitive activation of multiple candidates during prediction. To understand how comprehenders combine linguistic context with memory to generate candidates, researchers have often used anticipatory looks to targets simply as a read-out of that process. However, the method involves explicitly presenting candidates through a visual display, which makes observation of the candidates that are specifically generated based on the linguistic context challenging. To take one example, in Borovsky et al. (2012), children looked toward an appropriate target object before it was mentioned in the linguistic input, indicating that they had anticipated the target based on the preceding context. The authors, however, also observed a greater proportion of looks toward an action-related distractor that was an inappropriate patient of the verb given the preceding agent but semantically associated with the verb. For instance, upon hearing, "*The pirate chases the,*" children initiated more looks toward the correct target, ship, but also looked toward the verb-related distractor, cat, more than the other distractors, bone and treasure box. Increased looks toward the verb-related distractor might suggest that the contextually inappropriate candidate was considered in parallel with the target appropriate candidate. However, given the design of the visual world paradigm, it is possible that the presence of inappropriate candidates on the visual display forced the comprehender to consider candidates they would not otherwise generate in the absence of the visual input. The cat may act as a competitor only because it is presented as a potential target. It is, therefore, difficult to observe the candidates that children generate based on context in absence of explicitly presented candidates in a visual world paradigm. A similar concern is relevant in EEG/ERP paradigms, where responses to presented target words are measured, rather than directly probing self-generated predictions in advance of seeing the target words.

Another limitation of using comprehension-based experimental methods like the visual world paradigm is that it restricts the type of data that can be used to probe whether there is parallel activation of multiple candidates. Participants can only look at one target object at a given time, and aggregating across many individual looks does not straightforwardly show evidence for simultaneously active candidates. For example, a recent study reported that, like adults, 5-to-6-year-old children entertain multiple possibilities for upcoming words in a visual world paradigm (Sommerfeld et al., 2023). In that study, upon hearing, "The father eats now the…," children directed looks toward pictures of objects that were potential candidates that fit the semantic constraints of the preceding verb (i.e, edible objects), and they were equally distributed across multiple possible targets when presented with multiple visual objects that fit the context (pretzel, pizza, sausage, waffle). The authors concluded that children, like adults, consider multiple possibilities when predicting upcoming words based on preceding context. These findings are entirely consistent with the sampling approach (Smith & Levy, 2011), where a single response is sampled from a probability distribution on each trial. Determining whether the equal proportion of looks toward the different objects on the screen reflects parallel activation of those candidates and active competition among them is less straightforward.

1.3 The current study

We took advantage of the specific linking hypotheses between measures and underlying cognitive mechanisms proposed by the race model (Staub et al., 2015) and used the speeded cloze paradigm to probe the predictive mechanisms used by children. One possibility is that children and adults share the same mechanism and that they generate similar candidates for a given context. In that case, we would expect the same kind of competitive dynamics to appear in children's speeded cloze RTs as are found in adults. An alternative possibility is that the underlying mechanism is shared between children and adults but that the candidates and their properties differ between the two groups. For example, children might struggle to generate multiple viable candidates on some trials, due to differences in their world knowledge, vocabulary, or memory access mechanisms. In this case, we might still find evidence for systematic timing profiles in children's cloze responses, but we might not find the same specific effects of competition between candidates. It is also possible that children do not use the same race mechanism as adults, and that the race mechanism develops later in life. For example, children might sample words from a probability distribution (Smith & Levy, 2011), which does not involve competition among simultaneously generated candidates but rather a one-time sampling from a prior subjective distribution. Finally, it is possible that children's cloze behavior varies significantly across individuals, making it difficult to generalize a common mechanism based on group-based measures. This could be either because children do not have a shared mechanism as a group, or because they do not have the same kind of distribution of candidates which they sample from or use to generate candidates. Response patterns then may not be accurate reflections of a shared mechanism, as they would depend on each child's knowledge of linguistic and world constraints and other individual differences.

We examined whether children's speeded cloze responses exhibited the same pattern reflecting competitive dynamics as found with adults and as predicted by the race model. The study was carried out as part of a partnership with the Language Science Station at Planet Word, a language-focused museum in Washington D.C.. The Language Science Station brings research and researchers into the museum, with the aim to conduct research in a way that is mutually beneficial for researchers, students, and for museum visitors. Students receive training in science communication and conducting research via a summer course, and they staff booths at the museum to run different studies and talk with museum visitors about research. Carrying out the study in a museum setting made it possible to recruit a diverse group of participants, including early-grade children. We present data from mostly older children, with the average age being 9 years old and the age range 4-12 years old. The unique experiment setting also prompted the design of a child-friendly gamified version of the speeded cloze paradigm to fit the purpose of the research at the museum.

To preview the findings, children's and adults' response rates and RTs revealed the hallmark patterns reflecting race dynamics in both children and adults, supporting a shared mechanism in generating real-time predictions. There were, however, differences between children and adults regarding the response time evidence supporting the race mechanism. We show through a comparison of different measures of sentence constraints and race model simulations that the children's speeded cloze patterns that diverged from adults' can nevertheless be captured through a race process, where the outcomes are largely affected by the properties of simultaneously activated candidates.

2. Method

2.1 Participants

Participants were native English-speaking school-aged children (*N* = 60; ages 4-12, mean 9) and adults (*N* = 152; ages 13-59, mean 32)[2]. All participants voluntarily participated in the experiment as part of their experience at the Planet Word museum. Participants filled out a language background questionnaire at the beginning of the experiment which was used to screen participants' language and age background. All participants were sufficiently proficient readers to carry out the experimental task.

2.2 Materials

The experimental stimuli consisted of 40 sentence fragments with varying sentence constraints. They were created by truncating sentences with a range of high and low cloze probabilities, taken from a sentence norming study (Block & Baldwin, 2010). Table 1 presents examples of a high-cloze and low-cloze item. The 40 items were divided into two presentation lists, such that each participant saw only one of the two lists. Each list also included the same 20 filler items used for another experiment not reported here, consisting of object relative clause structures with a manipulation of word order to test the use of argument roles in verb prediction. In total, each participant saw 40 items, 20 critical items and 20 fillers, presented in random order. The full set of stimuli can be found in the Appendix.

Table 1. Example experiment stimuli

Sentence fragment	**Example response**	**Normed cloze probability***
Water and sunshine help plants	grow	0.95
A large stone blocked the entrance to the	cave	0.36

*The normed cloze values are from Block & Baldwin (2010)

[2] Due to the recruitment process where museum visitors voluntarily participated in the experiment, it was not feasible to restrict the age group of participants in advance. Analyses conducted with a more limited age range of 18-30 for adults yielded identical results. Therefore, we report the adult findings with the full range of the participants who were above age 12.

2.3 Procedure

Gamification of the speeded cloze paradigm

We designed a gamified, child-friendly version of the speeded cloze task, where the setup for the study introduced the idea of a robot as an example of an auto-complete technology. Participants were told that they would play a language science game called, 'Race the Robot,' and the goal of the game was to silently read sentence fragments and say aloud a continuation as quickly as possible. This version of the speeded cloze paradigm had the benefit of ensuring that participants, including young children, were engaged in the experiment in the less controlled museum setting. Our replication of the key adult findings from Staub et al. (2015), which used a speeded cloze paradigm in a more typical lab setting, confirms that the museum setting and the modified speeded cloze paradigm in the current experiment did not elicit different predictive behavior. Although not reported here, we also carried out a follow-up experiment using the same gamified speeded cloze paradigm and identical stimuli with adult English speakers on Amazon Mechanical Turk and replicated the main findings. Since the same results were obtained whether participants carried out the task in a quiet home setting or in the noisy museum setting, it is unlikely that the museum setting affected the results. The details of the online study can be found in Lee et al. (2023).

Task instructions were provided in verbal and written form, with three interleaved practice trials. In the practice trials, after participants produced a cloze response, a picture of a robot and an example completion of the sentence fragment were presented (Figure 1), to which participants could compare their own responses. No particular judgment was given by the experimenter about the responses that the participants produced, and no feedback was provided during the actual experiment. All participants provided verbal consent and completed a demographic survey before beginning the experiment. At the end of the experiment, the experimenter engaged in a conversation with participants about the purpose of the study and its potential relevance to everyday language use, as part of the learning experience at the museum. This was a distinctive feature of conducting research at a language museum and more broadly, promoting language science to the general public. Students received extensive coaching in how to hold these conversations, and museum visitors were highly engaged.

The experiment was administered using PCIbex (Zehr & Schwarz, 2018) on laptop computers. Each trial began with a '+' fixation mark presented for 1000 ms, followed by a sentence fragment presented word-by-word, with a 530 ms SOA (300 ms per word with a 230 ms blank screen in between). After the final word, a '?????' sign appeared, during which participants produced a completion. The screen automatically moved onto the next trial after a three-second time limit. Participants' responses were recorded using a directional microphone that was robust to surrounding noise in the museum. Each experiment session took about 10 minutes to complete, including consent and practice.

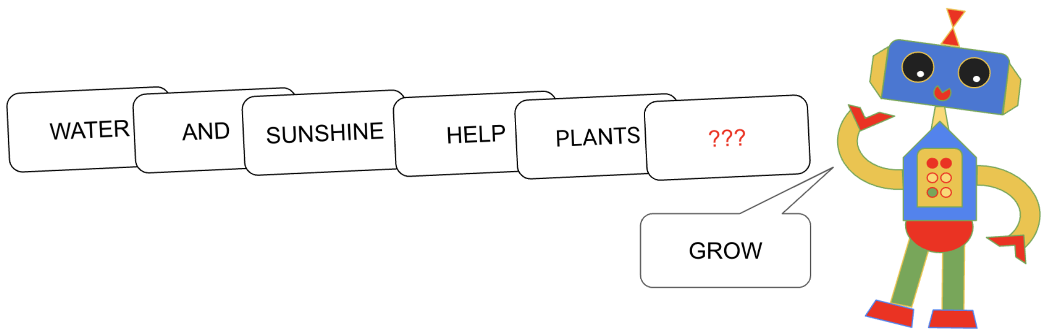

Figure 1. Illustration of the gamified speeded cloze experiment 'Race the Robot'

2.4 Analysis

2.4.1 Processing of speech data

The collected audio files were pre-processed by obtaining automatic transcriptions using Google Cloud Speech-to-Text API and by detecting automatic speech onset times using Chronset (Roux et al., 2017). The transcriptions and onset times were then manually checked and adjusted using Praat (Boersma, 2001). The review of transcriptions was carried out by the authors of this study. Codes that identified the experimental items were removed during this process, in order to avoid introducing biases in the manual review of transcriptions and onset times. Responses that were unidentifiable or that exceeded the 3-second time limit were removed from subsequent analyses.

2.4.2 Response time analyses

Cloze probabilities of responses were obtained by dividing the count of each response to a given sentence fragment by the total number of responses to that fragment, separately for children and adults. The response with the highest cloze probability for each item for each group was labeled as the item’s modal response, indicating the most frequently produced response for the item. Children and adults did not always have the same modal response. Following Staub et al. (2015), and as standard practice in the literature, we initially used the cloze probability of the modal response as a measure of ‘sentence constraint,’ or how constraining the particular item is. However, in subsequent analyses we will show that this is not the ideal measure and propose that mean response latencies more accurately reflect sentence constraint.

Statistical tests for the main effects of interest were conducted through linear mixed-effects models using statsmodels (Seabold & Perktold, 2010) in Python and logistic mixed-effects models for response rates using the lme4 library in R (Bates et al., 2015). All of the models initially included a maximal random effects structure for subjects, with the subject intercept and by-subject slopes for each of the fixed effects, and then were simplified until the model converged (Barr et al., 2013).

First, a model was built to examine the effect of cloze probability on RT, which included cloze probability, group, and their interactions as fixed effects. A separate model replacing cloze probability with the modal/non-modal contrast as a fixed effect was constructed to examine whether RTs differed between modal and non-modal responses (high- vs. low-cloze responses) at all levels of sentence constraint. Sentence constraint was represented by the modal cloze probability of each item, where modal cloze probability refers to the cloze probability of the response that was produced most frequently for an item. Finally, a model including only the non-modal responses was constructed, with cloze probability, sentence constraint, group, and their interactions as fixed effects, to examine the effect of sentence constraint on RT and whether it was consistent across responses with different cloze probabilities. Further details of the analyses, as well as follow-up analyses, are presented in the Results section.

2.4.3 Race model simulations

Analyses of the speeded cloze data were augmented with race simulations modeling different contexts in order to show how the observed empirical patterns can be derived from a race process. Further details of the simulations are provided along with the results in the next section.

3. Results

We first present the two key patterns found in children's and adults' speeded cloze responses, followed by an interim summary relating the observed findings to the race model. Then, we show through race simulations our proposed explanations for the observed child and adult differences.

3.1 The relationship between cloze probability and RT: *"Faster candidates, more frequent wins"*

Figure 2 presents the mean RTs of children's and adults' cloze responses plotted against cloze probabilities of the responses, where each point on the plot represents an individual response-context pairing. Children had longer response latencies than adults overall, with approximately a 200 ms delay (adults' mean RT: 589 ms; children's mean RT: 830 ms), but both groups showed the same relationship between cloze probability and RT, such that responses became faster as cloze probability increased.

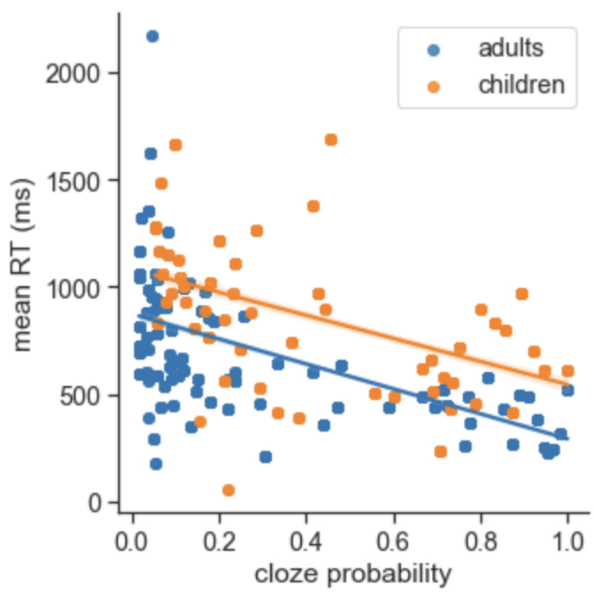

Figure 2. Relationship between cloze probability and RT of individual responses

The statistical model confirmed this pattern; there was a significant main effect of group ($\beta = .21$, $SE = .05$, $p < .001$), indicating the overall delay in children's RTs compared to adults', and a main effect of cloze probability ($\beta = -.59$, $SE = .03$, $p < .001$) and no interaction with group

($p > .05$), indicating that both children and adults produced high-cloze responses faster than low-cloze responses, and the size of the effect was similar between the two groups.

The correlation between cloze probability and response RT, where high-cloze responses were produced faster than low-cloze responses, can be explained by a race process, as discussed in detail in the Introduction. Faster candidates have a better chance of reaching the activation threshold first and winning the race, which means they should be produced more often than slower candidates. This predicts that responses with high cloze probability should have faster RTs, a prediction that is borne out in the speeded cloze data with both children and adults. However, this pattern could also be consistent with the sampling account, assuming that the strength of fit between a context and a candidate directly impacts the speed of the response in the same way that it affects the probability of sampling.

3.2 The relationship between sentence constraint and RT: *"Faster competitors, faster wins"*

Based on the race model, we expected to find a relationship between sentence constraint and RT, where highly constraining contexts should yield faster responses, at all levels of cloze probability. A candidate in a race must be faster than its competitors to win and get produced, and a candidate must be faster to win a more competitive race in a highly constraining context than to win a race with weaker competitors, due to the competitive dynamics of the race process.

Figure 3 shows the difference between the speed of non-modal responses produced in high versus low constraint items, with responses grouped according to their cloze probabilities. When responses were matched in terms of cloze probability, the responses were produced more quickly in highly constraining contexts than in less constraining contexts. For adults, this pattern was true across all cloze probability ranges. Children showed a similar relationship between non-modal response time and contextual constraint, except in the case of low cloze responses, where contextual constraint did not affect RT (the yellow and gray lines in the children's plot converge at low cloze probabilities).

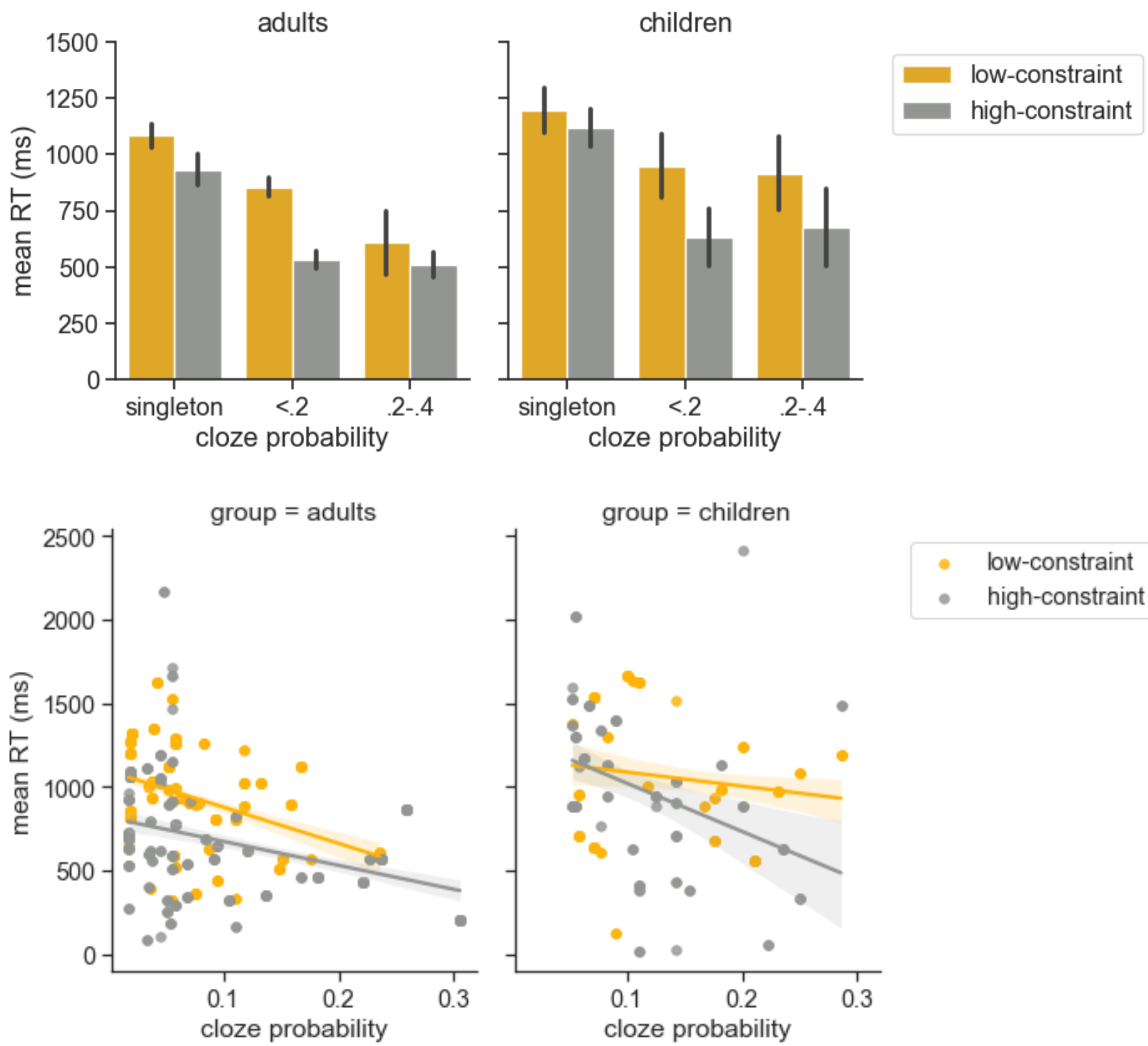

Figure 3. Relationship between sentence constraint and RT in high- and low-constraint items. Error bars represent standard error of the mean.

The statistical model revealed a main effect of cloze probability ($\beta = -1.90$, $SE = .95$, $p < .05$), a main effect of sentence constraint ($\beta = -.49$, $SE = .17$, $p < .01$), and, importantly, significant interactions with group, including an interaction between cloze probability and group ($\beta = 4.68$, $SE = 1.91$, $p < .05$), sentence constraint and group ($\beta = 1.23$, $SE = .44$, $p < .01$), and a three-way interaction among cloze probability, sentence constraint, and group ($\beta = -10.22$, $SE = 4.22$, $p < .05$). Post-hoc analyses for each of the groups revealed the following: adults' responses showed a main effect of cloze probability ($\beta = -1.87$, $SE = .85$, $p < .05$) and sentence constraint ($\beta = -.48$, $SE = .15$, $p < .01$) with no interaction between the effects ($p > .05$), indicating that higher cloze probability and higher sentence constraint independently contributed to faster RTs, replicating the pattern found with adults in Staub et al.'s (2015) experiments.

Conversely, children's responses did not show any significant main effects but showed an interaction between cloze probability and sentence constraint (β = -10.44, SE = 4.57, $p < .05$), where sentence constraint affected high-cloze non-modal responses more than low-cloze responses. Children's low-cloze responses were relatively slow regardless of how constraining the context was, and their responses were relatively slow in less constraining contexts, regardless of cloze probability. This pattern diverges from adults, and it does not naturally follow from the competitive dynamics predicted by the race model.

In sum, the analysis of sentence constraint revealed a contrast between children and adults. Adults consistently showed faster RTs for responses produced in highly constraining contexts, at all levels of cloze probability. This naturally follows from a race process where candidates have to be faster to win a race with stronger competitors than to win a less competitive race with weaker competitors. Children did not show the same relationship between constraint and RT as adults. In the following section, we address the divergent pattern in children by re-evaluating the measure of sentence constraint and by examining specific items where children and adults diverged the most. We show that i) the race model makes correct predictions regarding children's behavior once we use a more appropriate measure to classify sentence contexts according to the model's assumptions, and ii) the comparison between the different measures of sentence constraint is informative about potential differences between children's and adults' predictive processing.

3.3 Revisiting the race model predictions in relation to speeded cloze RTs

According to the race model, the average winning time of a candidate depends on the relative strength of its competitors. A race with strong competitors will generally produce fast winning times, regardless of which candidate wins the race, since a winner must be fast enough to beat the other competitors. However, a fundamental assumption of the model is that, during prediction, multiple lexical candidates are initially generated based on context, which then compete with each other toward a threshold. If only a few candidates are initially generated, this should affect the subsequent race process, including the competitive dynamics that yield the RT patterns that are observed in speeded cloze responses.

A closer examination of children's responses for specific items shows that children may, in fact, have fewer candidates relative to adults for some contexts. An example item that elicited the most divergent patterns between children and adults is presented in (1), with the list of distinct responses and the mean RTs of all responses produced for the item, by children and adults. Modal responses are indicated in bold.

(1) ***During autumn the air is crisp and…***				
	Responses	Modal cloze probability	Mean RT	Normalized* mean RT
adults	{*chilly, clean, clear,* **cold**, *cool, cream, dry, fresh, humid, light, ripe*}	0.41	695 ms	1348 ms
children	{***cold***, *sunny*}	0.80	1205 ms	1784 ms
(2) ***When she got out of the car she closed the…***				
	Responses	Modal cloze probability	Mean RT	Normalized* mean RT
adults	{*car,* ***door***}	0.98	324 ms	628 ms
children	{***door***, *window*}	0.88	472 ms	698 ms

*Normalized within each group, by subtracting the mean and dividing by the standard deviation.

For the context in (1), the modal response is the same in children and adults, but a diverse set of alternative responses in adults is replaced with a single alternative response in children. This suggests that children simply do not have the range of knowledge (or do not have the memory access abilities) that allows adults to generate so many different responses, and for that reason, some of the children's responses are notably slower, as reflected in the slow mean RT. These patterns contrast with the way children and adults responded to items like (2), which yielded comparable modal cloze probability, more similar mean RTs, and the same number of unique responses between the two groups. Items like (1) suggest that children and adults have different candidate profiles for some contexts, where children have fewer candidates than adults, which may lead to distinct speeded cloze RT patterns. This observation is further supported by the difference in the average number of unique responses produced for a context between the two groups, which was nine words for adults and five words for children.

The present analysis demonstrates a critical limitation of the standard analysis that represents sentence constraint using modal cloze probability. Doing so classifies sentence contexts like (1) and (2) into the same, high-constraint contexts for children, despite the noted differences in the candidate profiles. Moreover, it treats context (1) as more constraining for children than for adults, which is true under the standard definition of contextual constraint. However, to the extent that contextual constraint is understood to entail easy access to consistent responses, it is misleading. This motivated our choice to use an alternative measure that could more directly represent the overall strength of competitors in a given context, specifically the mean RT of all other responses produced for the context. For a particular candidate, a fast mean RT for the other responses indicates strong competitors, and hence, a more competitive race, which should yield a faster winning time for the candidate, compared to a situation where the other responses are overall slow, indicating weak competitors. We conducted an additional analysis of children's and adults' speeded cloze RTs using this alternative measure of sentence constraint and tested whether the empirical patterns naturally follow from a competitive race process, as proposed by the race model. The overall strength of competitors should affect individual response RTs due to an underlying race mechanism in prediction.

3.3.1 The relationship between sentence constraint and RT - revisited

As described above, the strength of competitors can be represented by the average of all other responses' RTs, where a fast mean RT indicates strong competitors and a slow mean RT indicates weaker competitors. Based on the race model, we expected to see faster individual response RTs when competitor strength was strong and to see this relationship hold at all levels of cloze probability; a race with fast competitors will generally produce a fast winning time regardless of who the winner is, since a winner must be fast to beat the other strong competitors.

Figure 4 shows the relationship between competitor strength and the RTs of individual responses. Unlike what was observed with modal cloze probability, the new analysis showed that at all levels of cloze probability, and in children and adults alike, responses were produced more quickly in contexts where their competitors were produced more quickly.

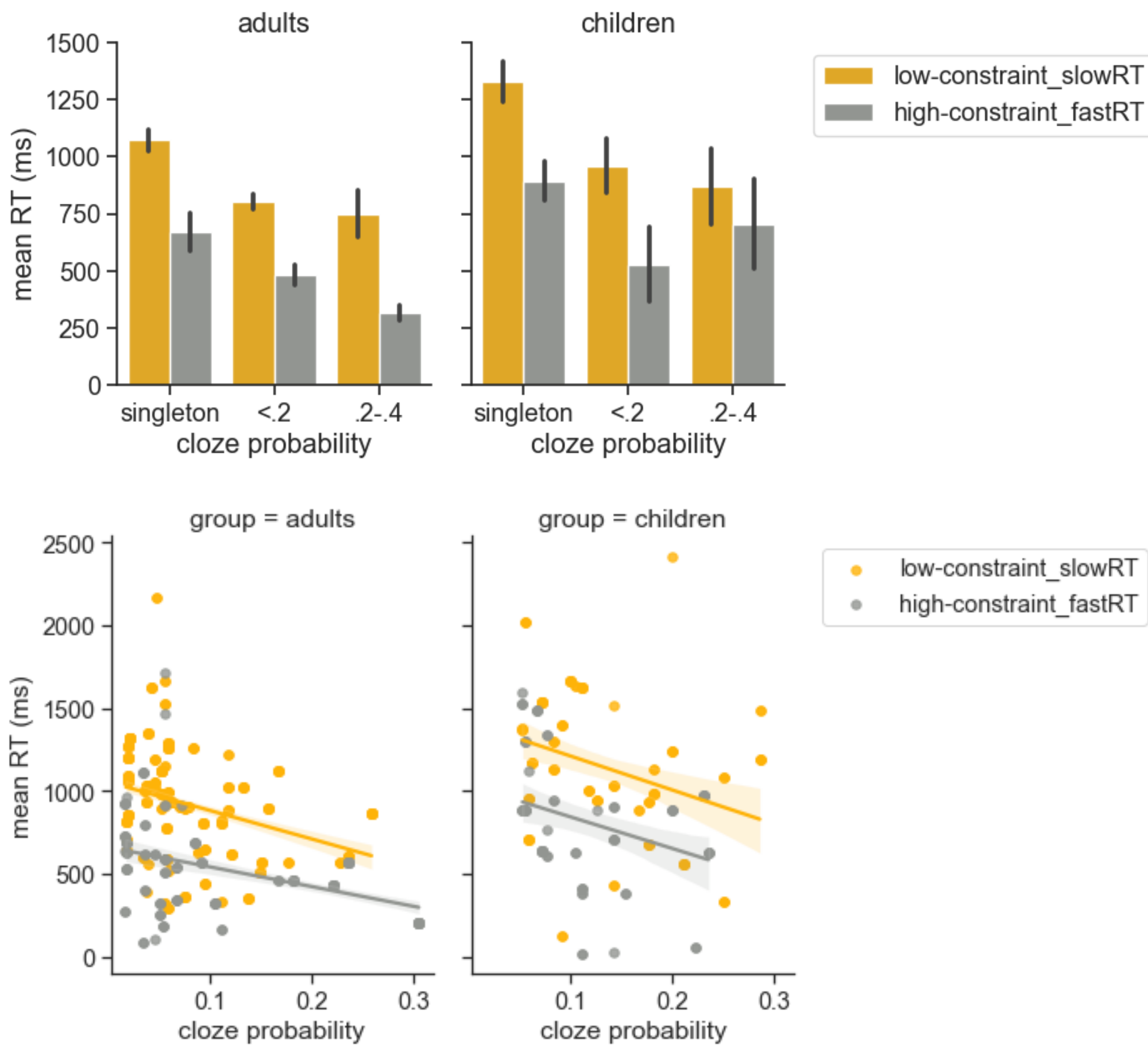

Figure 4. Relationship between sentence constraint and RT across different cloze probabilities. Error bars represent standard error of the mean.

A mixed effects model for target response RT, with the competitor strength (i.e., average RT of other candidates) as the fixed effect while maintaining the same random effects structure as before revealed a significant main effect of competitor strength ($\beta = .81$, $SE = .14$, $p < .001$) with no interaction with cloze probability or group (both $p > .05$). Children and adults produced responses faster when the context had faster candidates overall, compared to when the average RT of competitors was relatively slow.

The results show that having faster competitors yields faster responses at all levels of cloze probability, consistent with the race model's predictions, for both children and adults. In this respect, the speeded cloze responses reveal a shared predictive mechanism in children and adults which involves competition among multiple lexical candidates that race toward an activation threshold.

A remaining question is why the children did not show the same kind of relationship between sentence constraint and cloze RTs as found with adults. For adults, the two different ways of measuring sentence constraint yielded identical outcomes, whether sentence constraint was measured by modal cloze probability or by the average of competitors' RTs. Modal cloze probability characterizes a context, in terms of the strength of fit between the context and the set of possible continuations, through the existence of a dominant candidate (however fast), while mean competitor RT does so through the elicitation of fast candidates (however many there are). This indicates that for adults, a highly constraining context is typically a race with both a frequent winner and strong competitors, which leads to fast RTs overall. This was not true for children. The analysis with modal cloze probability revealed that children did not show a high-constraint and fast RT relationship at all levels of cloze probability, as the effect of sentence constraint on RT disappeared in the low-cloze responses.

We conducted race model simulations to model the kinds of contexts that could produce the children's speeded cloze response patterns which diverged from adults' (i.e., high modal cloze probability but slow average competitor RTs, or low modal cloze probability but fast average competitor RTs). We show that the children's patterns can also be captured by the same basic assumptions of the race model but with different candidate profiles compared to adults, where having fewer or slower candidates leads to the unique RT patterns in children.

3.4 Race model simulations

In the simulations, a race context is defined through a set of candidates, where each candidate has a mean finishing time and a distribution, which together characterize the relative strength of each candidate given a particular context. In a single trial of a race, a single finishing time is sampled from each candidate's distribution, and the means are then ranked to determine the candidate with the fastest finishing time, i.e., the cloze response. Multiple trials are simulated for each race context. In this way, the win proportions of each candidate and its average winning time across the multiple trials can be calculated and compared across different race contexts.

We first replicated Staub et al.'s (2015) simulations, which demonstrated that a highly constraining context yields faster finishing times for candidates at all levels of cloze probability. The default race context was defined by 10 candidates with each of their mean finishing times, evenly distributed with 10 ms increments from 955 ms to 1045 ms (i.e., 955 ms, 965 ms, 975 ms, …, 1045 ms; $N = 10$) with a set variance of 50 ms. Then, highly constraining contexts were created by changing the default candidate set in two different ways: i) increasing the difference between the means for different candidates to 20 ms instead of 10 ms, or ii) adding a candidate with an exceptionally fast mean finishing time (i.e., 915 ms, 955 ms, 965 ms, 975 ms, …, 1045 ms; $N = 11$). A total of 100,000 trials were simulated for each race context. They showed that highly constraining contexts yield higher win proportions for the modal response and faster finishing times for all candidates in the race compared to the default context. The reason for this is that, in both scenarios, the strongest competitor becomes harder to beat, and so the other candidates win the race only on their fastest trials.

In order to model races that could capture children's speeded cloze patterns, where high modal cloze probability did not always yield fast mean RT, we manipulated the competitors in Staub et al.'s highly constraining context in two different ways: i) by changing the competitors' mean finishing times, or ii) by changing the number of competitors. These manipulations were motivated by the observation of items where children notably produced fewer distinct responses than adults, as discussed in the previous section. First, in order to model the effect of competitor strength on race outcomes, we took Staub et al.'s highly constraining context with an exceptionally fast candidate and increased the speed of each candidate by 20 ms, except for the fastest candidate, which served as the target for comparison between the different contexts simulated. The resulting set of competitors in the new highly constraining context, therefore, had faster candidates (e.g., 915 ms, 935 ms, 945 ms, …, 1025 ms; $N = 11$) than in the original set. Additionally, a race with slower competitors was created by adding 20 ms to each of the original candidate's mean finishing times (e.g., 915 ms, 975 ms, 985 ms, …, 1065 ms; $N = 11$). A total of 100,000 trials were simulated for two race contexts, and the results are plotted in Figure 5, along with Staub et al.'s default and highly constraining contexts.

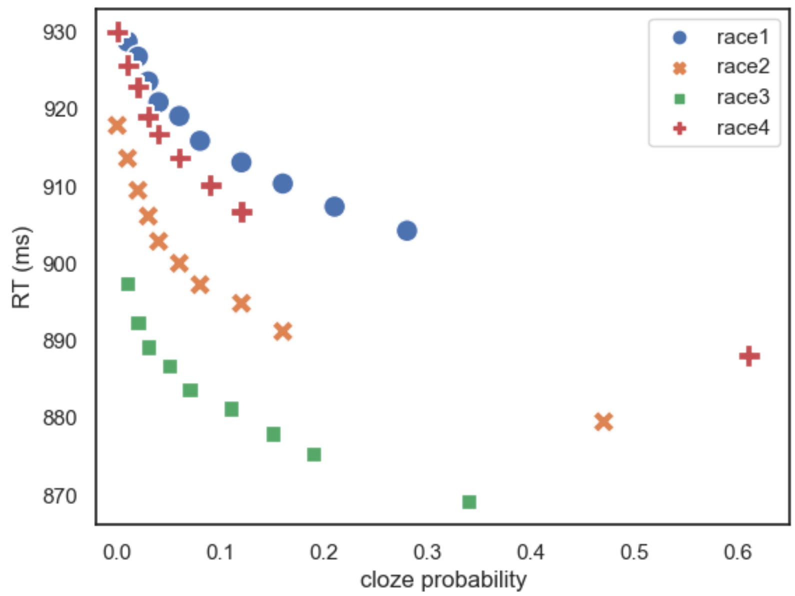

Figure 5. Results of race simulations modeling contexts with different speed of competitors

Compared to the default context (Race 1), all of the highly constraining contexts (Race 2-4) yielded higher modal cloze probability, represented in the plot by the rightmost candidate in each race context. However, within the highly constraining contexts, there was variability in modal cloze probability depending on the speed of the competitors. Note that this modal candidate has exactly the same underlying finishing time distribution in each of these three simulations (Races 2-4). Compared to Staub et al.'s Race 2, the race representing a highly constraining context with faster competitors yielded faster finishing times overall but lower modal cloze probability (Race 3), while the race with slower competitors yielded slow finishing times overall but higher modal cloze probability (Race 4). The results demonstrate that the strength of competitors can lead to variation in modal cloze probability, which could explain children's speeded cloze patterns: a candidate can win more often despite being slow if its competitors are weak, hence producing a high modal cloze probability and slow mean RT pattern.

We also simulated race contexts that had half or double the number of competitors in Staub et al.'s highly constraining context with the exceptionally fast candidate. Again, 100,000 trials were simulated for each of the races, and the results are plotted in Figure 6.

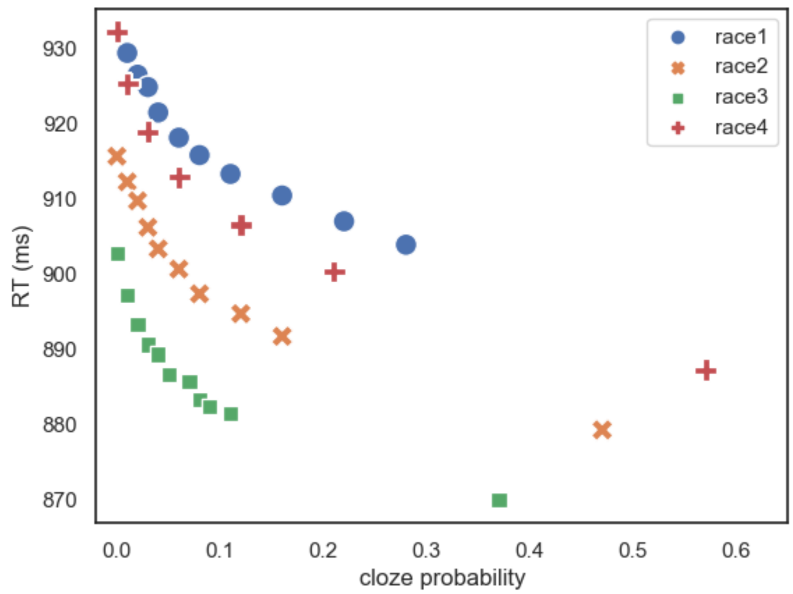

Figure 6. Results of race simulations modeling contexts with different number of competitors

The results replicated the pattern observed in the previous simulation that manipulated the speed of the competitors. The race with more competitors elicited faster RTs but lower modal cloze probability (Race 3), while the race with fewer competitors produced slower RTs but higher modal cloze probability (Race 4).

The simulations combined show that the speed of competitors and the number of competitors can influence race outcomes, either of which yields the pattern observed in children's speeded cloze responses. Staub et al.'s simulations demonstrated that a highly constraining context could be a race where there is an exceptionally fast candidate, which yields faster overall winning times, due to a competitive race process. Our simulation results suggest that for children, the same kind of competitive dynamics occur and that a highly constraining context for children can similarly be a race with a fast candidate but with weaker (slower or fewer) competitors in the race. This results in a pattern that diverges from adults: highly constraining contexts sometimes produce slow RTs, particularly in low-cloze responses that rarely win the race.

4. Discussion

The current study examined children's self-generated speeded cloze productions and whether they show the same hallmark patterns as adults, reflecting an underlying competitive dynamic, involving multiple lexical candidates racing toward an activation threshold to get produced (Staub et al., 2015). We designed a gamified version of the speeded cloze paradigm and collected responses from a large sample of children and adults, taking advantage of visitor interest in a language-focused museum. The analysis of response rates and RTs revealed that while children's responses were slower than adults', they showed both key patterns that suggest the same competitive race mechanism in generating predictions in real-time. For adults, i) high-cloze responses were produced faster than low-cloze responses, at all levels of sentence constraint, and ii) highly constraining contexts produced faster responses than low constraining contexts, at all levels of cloze probability, supporting the race model. Children showed the same cloze and RT relationship but diverged from adults regarding the effect of sentence constraint on RT. However, we showed that this was due to an inappropriate measure of sentence constraint, and using an alternative measure resulted in identical patterns between children and adults, supporting parallel, competitive activation of multiple candidates based on context. Through race model simulations, we showed that the children's unique speeded cloze pattern could be a result of having slower or fewer competitors for a dominant candidate given a highly constraining context. We address these key findings in more detail below.

4.1 A race mechanism in children's real-time generation of next-word predictions

We expected two key patterns to appear in children's speeded cloze response RTs if children used a race-like process to generate candidates in a speeded cloze paradigm like adults. One was the relationship between cloze probability and RT, where high-cloze responses should be faster than low-cloze probability, regardless of how constraining the context is. This is because as candidates accumulate activation toward a threshold, faster candidates have a greater chance of reaching the threshold first and getting produced than slower candidates, resulting in faster responses for higher cloze probability. We found this relationship in children as well as adults, where high-cloze responses were overall faster than low-cloze responses, regardless of how constraining the context is. This pattern is consistent with a race-like process in generating predictions online which is shared between children and adults.

One might argue that the relationship between cloze probability and RT is also consistent with an alternative mechanism, where cloze responses are sampled from a probability distribution (Smith & Levy, 2011). The original sampling account, as discussed in the

Introduction, does not make specific timing predictions, as it simply involves a one-time sampling process that was proposed to explain offline cloze probability. However, one could extend the theory to predict a meaningful relationship between cloze probability and RT, where it would have the same kind of relationship that is proposed to hold between cloze probability and subjective probability distributions. If people generate cloze responses by sampling from a distribution that reflects how good of a fit a candidate is given a context, that goodness of fit could be directly reflected in response times, in the same way as it is reflected in cloze probability. Stronger candidates would be associated with higher cloze probability and faster response times. In this way, the relationship between cloze probability and RT observed in both children and adults in the current study would be compatible with a shared mechanism where cloze responses are generated by sampling from prior distributions.

While the relationship between cloze probability and response RT might be captured by the two alternative mechanisms, the second key finding, where highly constraining contexts yielded faster responses than less constraining contexts, supports the race model over a sampling account. According to the race model (Staub et al., 2015), the relationship between sentence constraint and RT is a consequence of a competitive race process where a candidate must be faster than its competitors to get produced, so that even non-dominant candidates must be faster to beat stronger competitors in a highly constraining context. The adults' speeded cloze responses in the present study showed the expected pattern: highly constraining contexts produced faster responses than less constraining contexts, and this effect did not interact with cloze probability, replicating Staub et al.'s (2015) findings and confirming the model's predictions. The children's responses, however, diverged from those of adults, crucially, when sentence constraint was measured by modal cloze probability. However, using a more appropriate measure of sentence constraint, specifically the mean RTs of all other responses produced for a context, excluding the RT of the word itself, revealed the expected relationship between constraint and RT in both children and adults. Children's speeded cloze response patterns are, therefore, consistent with a race-like process in generating next-word predictions, where multiple candidates compete with each other to reach a threshold and get produced. On the other hand, the constraint-RT relationship is difficult to explain with a sampling account, as it is not straightforward how the relative strength of competitors should affect the RT and cloze probability of a candidate given a context.

We additionally ran race model simulations in order to examine what kind of candidate profiles would yield children's speeded cloze patterns that diverged from adults, where contexts with high modal cloze probability did not consistently yield fast mean RT and vice versa, low modal cloze probability did not always represent a context with slow responses. The simulation results showed that either reducing the number of competitors or increasing the time it takes for competitors to reach threshold reproduced the children's speeded cloze RT patterns. This suggests that children's highly constraining contexts may be contexts that have fewer or slower competitors for a dominant candidate compared to adults'. Importantly, the results suggest that the divergence between children and adults can be explained by a difference in candidate profiles rather than by a fundamentally different predictive mechanism. Children's response patterns naturally follow from a race mechanism proposed to explain adults' speeded cloze behavior.

4.2 Theoretical contributions and implications

The main contributions of the current study to the existing literature on children's predictive processing are twofold. It offers an effective methodology to study predictions in children, and it helps to answer theoretical questions about the underlying cognitive mechanisms involved in children's predictive behavior.

The speeded cloze paradigm used in this study addresses some shortcomings of previous measures used to study children's predictive processing. Much of the literature has focused on whether or not children show predictive behavior like adults in comprehension measures like eye-movements in a visual world paradigm. Exhibiting adult-like patterns in these tasks may suggest that children can analyze the linguistic input and use it to correctly select or respond to an appropriate target given a set of presented options. However, since these experiment designs require the presentation of visual objects or linguistic input (in EEG/ERPs) which serve as the target or distractors to which responses are measured, it is difficult to directly probe whether children generate multiple candidates in parallel based on context alone. Our findings show that the speeded cloze paradigm offers a more direct way of examining children's predictive processes, where response latencies offer informative evidence in support of a competitive race-like mechanism in generating predictions online. We are currently conducting a follow-up study using a reading-free version with auditory context presentation, which makes it possible to test younger children.

The fact that children's speeded cloze responses in the current study could be captured through the same race process proposed to explain adults' predictive behavior suggests that the same mechanism is used from a young age rather than developed later in life. Although the child participants in the current study were older than the age group typically examined in previous studies, the findings suggest that the mechanism involving competition among multiple candidates is not unique to adults and may be shared with younger children. An overall slowdown in RTs at younger ages, for example, is unlikely to affect the main patterns observed with the older children and adults in the current study. The results challenge alternative potential theories for explaining how children generate speeded cloze responses, such as through a non-parallel activation of a single best candidate with no competition involved or by sampling candidates from a probability distribution (Smith & Levy, 2011), which cannot capture the systematic relationship between sentence constraint and RT observed with children as well as adults. Our findings can also provide new insights to theories that claim prediction is a driving factor of children's language development (i.e., error-based learning: Elman, 1990; Chang et al., 2006; Dell & Chang, 2014; Fazekas et al., 2020; Peter et al., 2015; see also Rabagliati et al., 2016; Lidz et al., 2017), and future work could investigate how a race mechanism in generating candidates during real-time prediction affects children's acquisition of new linguistic knowledge. Finally, the results speak to the broader literature on prediction (Altmann & Mirković, 2009; Van Petten & Luka, 2012; Dell & Chang, 2014; Huettig, 2015; Kuperberg & Jaeger, 2016; Pickering & Garrod, 2013; Pickering & Gambi, 2018; Federmeier, 2007, 2022) and support the context-driven parallel activation of multiple candidates for upcoming words in real-time prediction (Roland et al., 2012; Staub et al., 2015; Frisson et al., 2017; Ness & Meltzer-Asscher, 2021).

The current study additionally highlights the benefit of conducting research outside a lab setting where more active interactions with the participants and the general public was possible. The experiment setup at the language museum worked remarkably well. We were able to collect good spoken language data, the children and parents enjoyed participating in the gamified experiment, and the students who led the sessions on-site benefited from their training. The museum found that it enriched their visitors' experience. This opens useful avenues for testing large numbers of participants outside a lab setting. Importantly, it depended on very careful attention to designing the study and training the student researchers, so as to create a positive experience.

5. Conclusion

We used a speeded cloze paradigm to probe the underlying predictive mechanisms in school-aged children as well as adults. We examined whether children's speeded cloze responses showed adult-like patterns that reflect a competitive race dynamic, where multiple candidates accumulate activation based on context and compete toward a threshold (Staub et al., 2015). We recruited child and adult visitors at a language-focused museum and administered a gamified version of the speeded cloze task, collecting self-generated responses and RTs. The analyses of the speeded cloze data augmented with computational race simulations showed that children's response patterns can also be captured by a race mechanism in generating candidates for upcoming words during real-time sentence processing, although for some highly constraining contexts, children may generate fewer competitors or weaker competitors for a dominant candidate compared to adults, leading to RT patterns that diverge from adults.

Appendix: Experiment stimuli

Sentence fragment	**Example response**	**Normed cloze probability (Block & Baldwin, 2010)**
She could tell he was mad by the tone of his	voice	0.99
She went to the bakery for a loaf of	bread	0.98
The dentist says brushing your teeth twice a	day	0.98
After hitting the iceberg the ship began to	sink	0.96
Water and sunshine help plants	grow	0.95
She wore a colorful scarf around her	neck	0.95
When she got out of the car she closed the	door	0.95
The children went outside to	play	0.94
After dinner he washed his hands with	soap	0.94
I roasted the marshmallow over the	fire	0.89
The janitor accidentally spilled some water on the	floor	0.88
Charles dunked the basketball through the	hoop	0.87
At Jessie’s birthday party they ate a delicious	cake	0.95
There were no extra seats so she sat on the	floor	0.83
Derek's feet were cold, so he put on some	socks	0.82
After winning the carnival game, Tim received a	prize	0.82
The man happily sat down in the comfortable	chair	0.82
Maggie kept her wallet and keys inside her	purse	0.77
Pam did not have any clothes to	wear	0.74

The pinecone fell and hit Terry in the	head	0.74
She went to bed because she was	tired	0.72
On her birthday she excitedly opened the	gifts	0.71
The deer ran out of the woods and across the	road	0.67
Betty did not laugh when she heard the	joke	0.65
The child could not sleep without his stuffed	bear	0.63
She followed the recipe correctly to cook the	meal	0.56
They were startled by the sudden	noise	0.5
The milk sat out so long it began to	spoil	0.49
After being in the cold Jake's fingers were	numb	0.4
A large stone blocked the entrance to the	cave	0.36
During autumn the air is crisp and	cold	0.35
Jeffrey didn't get the question right but his answer was	good	0.23
Zack went to the supermarket and bought a	NA	NA
Bob wanted a snack so he ate a	NA	NA
Olivia opened the cabinet to take out the	NA	NA
Liam worked really hard to finish the	NA	NA
Emma peeled off the sticker and put it on the	NA	NA
Henry was holding the	NA	NA
Johnny was looking for his	NA	NA
Sophia was happy when she saw the	NA	NA

References

Altmann, G. T., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition, 73*(3), 247-264.

Altmann, G. T., & Mirković, J. (2009). Incrementality and prediction in human sentence processing. *Cognitive Science, 33*(4), 583-609.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255-278.

Bates D, Mächler M, Bolker B, Walker S (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1–48.

Boersma, Paul (2001). Praat, a system for doing phonetics by computer. *Glot International 5:9/10*, 341-345.

Borovsky, A., Elman, J. L., & Fernald, A. (2012). Knowing a lot for one's age: Vocabulary skill and not age is associated with anticipatory incremental sentence interpretation in children and adults. *Journal of Experimental Child Psychology, 112*(4), 417-436.

Brown, V. A., Fox, N. P., & Strand, J. F. (2022). "Where are the... Fixations?": Grammatical number cues guide anticipatory fixations to upcoming referents and reduce lexical competition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 48*(5), 643.

Dell, G. S., & Chang, F. (2014). The P-chain: Relating sentence production and its disorders to comprehension and acquisition. *Philosophical Transactions of the Royal Society B: Biological Sciences, 369*(1634), 20120394.

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience, 8*(8), 1117-1121.

Ehrlich, S. F., & Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. *Journal of Verbal Learning and Verbal Behavior, 20*(6), 641-655.

Fazekas, J., Jessop, A., Pine, J., & Rowland, C. (2020). Do children learn from their prediction mistakes? A registered report evaluating error-based theories of language acquisition. *Royal Society Open Science, 7*(11), 180877.

Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology, 44*(4), 491-505.

Federmeier, K. D. (2022). Connecting and considering: Electrophysiology provides insights into comprehension. *Psychophysiology, 59*(1), e13940.

Federmeier, K. D., & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language, 41*(4), 469-495.

Frisson, S., Harvey, D. R., & Staub, A. (2017). No prediction error cost in reading: Evidence from eye movements. *Journal of Memory and Language*, *95*, 200-214.

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

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.

Gertner, Y., Fisher, C., & Eisengart, J. (2006). Learning words and rules: Abstract knowledge of word order in early sentence comprehension. *Psychological Science, 17*(8), 684-691.

Hoeks, J. C., Stowe, L. A., & Doedens, G. (2004). Seeing words in context: the interaction of lexical and sentence level information during reading. *Cognitive Brain Research, 19*(1), 59-73.

Huettig, F. (2015). Four central questions about prediction in language processing. *Brain Research, 1626*, 118-135.

Kamide, Y., Altmann, G. T., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language, 49*(1), 133-156.

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension?. *Language, Cognition and Neuroscience, 31*(1), 32-59.

Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences, 4*(12), 463-470.

Lee, E., Howitt, K., Dixon, L., Ness, T., Nakamura, M., & Phillips, C. (2023). Alignment between adult and child predictive processing dynamics: Evidence from a gamified cloze study in a museum. Poster presented at *The 36th Annual Conference on Human Sentence Processing*, University of Pittsburgh, Pennsylvania, USA.

Lidz, J., White, A. S., & Baier, R. (2017). The role of incremental parsing in syntactically conditioned word learning. *Cognitive Psychology, 97*, 62-78.

Nakamura, M. (2023). *Generating and measuring predictions in language processing* [Doctoral dissertation, University of Maryland].

Author's Last Name, First Initial. Middle Initial. (Year). *Title of dissertation/thesis* (Publication No. #) [Dissertation type, University]. Database. URL

Mani, N., & Huettig, F. (2012). Prediction during language processing is a piece of cake—But only for skilled producers. *Journal of Experimental Psychology: Human Perception and Performance, 38*(4), 843.

Mani, N., & Plunkett, K. (2010). In the infant's mind's ear: Evidence for implicit naming in 18-month-olds. *Psychological Science, 21*(7), 908-913.

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

Ness, T., & Meltzer-Asscher, A. (2021). Love thy neighbor: Facilitation and inhibition in the competition between parallel predictions. *Cognition, 207*, 104509.

Özge, D., Küntay, A., & Snedeker, J. (2019). Why wait for the verb? Turkish speaking children use case markers for incremental language comprehension. *Cognition, 183*, 152-180.

Özge, D., Kornfilt, J., Maquate, K., Küntay, A. C., & Snedeker, J. (2022). German-speaking children use sentence-initial case marking for predictive language processing at age four. *Cognition, 221*, 104988.

Pickering, M. J., & Gambi, C. (2018). Predicting while comprehending language: A theory and review. *Psychological Bulletin, 144*(10), 1002.

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences, 36*(4), 329-347.

Roland, D., Yun, H., Koenig, J. P., & Mauner, G. (2012). Semantic similarity, predictability, and models of sentence processing. *Cognition, 122*(3), 267-279.

Seabold, S., & Perktold, J. (2010). statsmodels: Econometric and statistical modeling with python. *In 9th Python in Science Conference*.

Sommerfeld, L., Staudte, M., Mani, N., & Kray, J. (2023). Even young children make multiple predictions in the complex visual world. *Journal of Experimental Child Psychology, 235*, 105690.

Smith, N., & Levy, R. (2011). Cloze but no cigar: The complex relationship between cloze, corpus, and subjective probabilities in language processing. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 33, No. 33).

Smolík, F., & Bláhová, V. (2022). Here come the nouns: Czech two-year-olds use verb number endings to predict sentence subjects. *Cognition, 219*, 104964.

Staub, A. (2015). The effect of lexical predictability on eye movements in reading: Critical review and theoretical interpretation. *Language and Linguistics Compass, 9*(8), 311-327.

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.

Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology, 83*(2), 176-190.

Zehr, J., & Schwarz, F. (2018). PennController for Internet Based Experiments (IBEX).