

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository: <https://orca.cardiff.ac.uk/id/eprint/116135/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Lindsay, Laura, Gambi, Chiara and Rabagliati, Hugh 2019. Preschoolers optimize the timing of their conversational turns through flexible coordination of language comprehension and production. *Psychological Science* 30 (4) , pp. 504-515. 10.1177/0956797618822802

Publishers page: <https://doi.org/10.1177/0956797618822802>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See <http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.



Preschoolers optimize the timing of their conversational turns through flexible coordination
of language comprehension and production.

Introduction and Discussion: 2000

Total (excluding references and abstract): 5826

Laura Lindsay*

University of Edinburgh

Chiara Gambi*

University of Edinburgh and Cardiff University

Hugh Rabagliati

University of Edinburgh

Address for correspondence:

Chiara Gambi

School of Psychology

70 Park Place

Cardiff University

CF10 3AT Cardiff, U.K.

Email: GambiC@cardiff.ac.uk

*Joint first authorship

Abstract

Conversation is the natural setting for language learning and use, and a key property of conversation is the smooth taking of turns. In adult conversations, delays between turns are minimal (typically 200ms or less), because listeners display a striking ability to predict what their partner will say, and formulate a response before their partner's turn ends. Here, we test how this ability to coordinate comprehension and production develops in preschool children. In an interactive paradigm, one-hundred-and-six 3-to-5-year-olds (and forty-eight adults) responded to questions that varied in predictability, but were controlled for linguistic complexity. Using a novel distributional approach to data analysis, we show that when children can predict a question's ending, they leave shorter gaps before responding, suggesting that they, like adults, can optimize the timing of their conversational turns. In line with a recent ethological theory of turn-taking, this early competency helps explain how conversational contexts support language development.

Keywords: Turn-Taking; Prediction; Response Planning; Development; Conversation

Preschoolers optimize the timing of their conversational turns through flexible coordination of language comprehension and production.

For adults, conversation is the most natural form of social interaction, and a key characteristic of conversation is the smooth taking of turns. Timely turn-taking appears to be universal (Levinson, 2006): Across diverse languages and cultures, conversational partners often switch turns, rarely speak over one another, and typically leave only the smallest possible silent gap between turns (Stivers et al., 2009). The brevity of such silent gaps (often less than 200ms) is particularly striking, because linguistic formulation takes time: Preparing to name a picture, for instance, takes at least 600ms (Indefrey & Levelt, 2004). Thus, speakers must begin formulating their responses even while their partner is still talking, and this act of formulation must itself be based on a prediction about what their partner was going to say (Bögels, Magyari, & Levinson, 2015; Corps, Crossley, Gambi, & Pickering, 2018; Levinson, 2016; Magyari, Bastiaansen, de Ruiter, & Levinson, 2014; Riest, Jorshick, & de Ruiter, 2015; Sacks, Schegloff, & Jefferson, 1974; though see Sjerps & Meyer, 2015). In other words, the optimal timing of conversational turns depends upon an ability to coordinate language comprehension (prediction) and production (formulation).

In recent work, Levinson (2006; 2016) has argued that fluent turn-taking may be “part of our ethology” (Levinson, 2016; p. 10), that is to say, an evolved adaptive behaviour that facilitates communication, and a skill upon which the development of other linguistic and cognitive skills might depend (Clark, 2007; Gelman S., 2009; Hirsh-Pasek et al., 2015; Zimmerman et al., 2009). Based on this, he proposes that the ability to engage in fluent turn-taking, including the ability to coordinate prediction and formulation, should emerge early in development. And indeed, observational studies show that pre-linguistic infants engage in

“proto-conversational” interactions with their caregivers (Jaffe, Beebe, Feldstein, Crown, & Jasnow, 2001; Murray & Trevarthen, 1986), and that these interactions are well-coordinated and well-timed (Hilbrink, Gattis, & Levinson, 2015).

However, as children begin to interact using language, their fluency at turn-taking declines, such that they leave longer gaps during linguistic turn-taking than during previous non-linguistic turn-taking (Hilbrink et al., 2015). One recent corpus analysis found that the median gap left by toddlers and preschoolers was greater than 600ms, i.e., hundreds of milliseconds longer than the typical gap left by adults (Casillas, Bobb, & Clark, 2016), while other work reported median gaps of greater than 1000ms even in 5-year-olds (Garvey & Berninger, 1981; Stivers, Sidnell, & Bergen, 2018). This developmental trajectory -- from well-timed non-linguistic interactions to less fluent linguistic interactions -- raises the possibility that preschoolers may in fact lack competence at flexibly coordinating linguistic prediction and formulation. For example, they may only start formulating a response when their partner is about to stop speaking (as Sjerps & Meyer, 2015, suggested for adults), which would also limit the overlap between comprehension and production, and thus make switching between tasks less taxing (Zelazo et al., 2003).

In response, Levinson (2016) has argued that the poor timing of children’s conversational turn-taking is not indicative of an underlying lack of competence, but instead results from “the challenge of cramming...complex linguistic material into brief turns” (Levinson, 2016; p.10; Casillas, et al., 2016). That is to say, children can flexibly coordinate prediction of what a conversational partner will say with early formulation of a response but, compared to adults, they find it hard to linguistically encode the ideas that they want to express, and thus respond with a delay. Consistent with this, eye-tracking studies suggest that children are skilled predictors. For instance, when one-year-olds observe two people conversing, they anticipatorily gaze to the next speaker as the turn changes (Casillas & Frank, 2017; see also

Keitel, Prinz, Friederici, von Hofsten, & Daum, 2013), while even two-year-olds can predict upcoming words when listening to simple sentences (Borovsky, Elman, & Fernald, 2012; Gambi, Pickering, & Rabagliati, 2016; Mani & Huettig, 2012; Rabagliati, Gambi, & Pickering, 2015). But importantly, this experimental work only assesses how children process language while passively listening.

To directly test the key claim that children can coordinate prediction and formulation, we developed a paradigm in which children were active contributors to an experimentally-controlled conversation, embedded in an interactive touch-screen maze game. Our design was inspired by the finding that adults leave shorter gaps after questions containing earlier-occurring informative material (e.g., *which character, also called 007, appears in the famous movies?* versus *which character from the famous movies is also called 007?*), suggesting that they prepare a response as soon as possible (Bögels, et al., 2015). We built on this demonstration to test whether children can coordinate prediction of a question with early preparation of their response.

In our paradigm, children conversed with an avatar (Peter Pan) as he played hide-and-seek with a parrot around a maze. When Peter Pan reached a fork in the maze, he encountered two familiar cartoon characters (e.g., Po and Boots, depicted schematically in Figure 1), and the parrot hid behind one. Peter Pan then asked the participant where the parrot was hiding, to which they responded verbally. We expected children to leave shorter conversational gaps when the informative content (the name of one of the cartoon characters) arrived earlier in the question (e.g., *Is Po hiding the parrot?*) than when it arrived later (e.g., *Is the parrot behind Po?*).

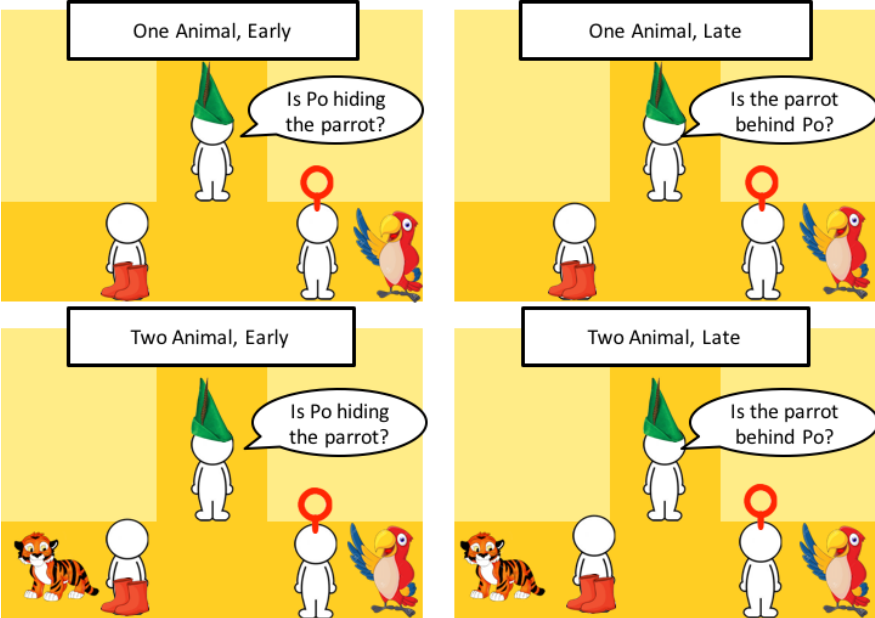
To account for how extraneous linguistic differences between the questions may affect response times, we created control mazes where Peter Pan chased and asked about two animals









(a parrot and a tiger; Figure 1). At each fork, the two animals hid behind different cartoon characters and Peter Pan asked where one had hidden (e.g., *Is Po hiding the parrot?* or *Is Po hiding the tiger?*). Thus, because Peter Pan could ask about either animal, participants could not accurately predict what he would say, regardless of whether the relevant information came early or late in the question. Crucially, if participants can use predictions to begin formulating a response before a question ends, then we expect an interaction between the information structure of the question and the type of maze: In one-animal mazes, participants should be faster to respond after early than late questions, and this difference in response times should be greater than in two-animal mazes, where prediction is not possible.



To test for the presence of this interaction in adults and preschoolers, we adopted a novel distributional data analysis method. Since the distribution of gaps in turn-taking is heavily skewed to the right (especially for children), standard statistical comparisons between means are problematic, as skew can mask true differences in response times, or induce spurious ones. We thus moved away from analysing the mean, and instead modelled how the full distribution of response times changed across conditions (Umlauf & Kneib, 2018). Our statistical model was based on the ex-Gaussian distribution, in which a Gaussian and exponential distribution are convolved. This provides an excellent fit to right-skewed data: The mean of the Gaussian captures shifts in modal response times, and the rate of the exponential captures changes to the weight of the right tail (Balota & Yap, 2011). Previously, developmental researchers have rarely used ex-Gaussian analyses, because such models are typically fit to individual participants, and thus require large numbers of observations. Here, however, we built on recent statistical advances (Umlauf & Kneib, 2018) to fit a hierarchical, or multi-level, ex-Gaussian model, where observations are partially pooled across participants (Gelman A. & Hill, 2007), resulting in a rich description of response dynamics during turn-taking.

RUNNING HEAD: PRESCHOOLERS' TURN-TAKING

Figure 1. Visual representation of conditions (top) and summary of predictions (bottom). Characters are depicted schematically here, but participants saw actual depictions of the characters when playing the game.



Maze Type	Information Structure	Characters Present	Question	Can predict animal?	Can prepare early?
One animal	Early	 	Is Po hiding the parrot?	Yes	✓
	Late	 	Is the parrot behind Po?	Yes	✗
Two animals	Early	 	Is Po hiding the parrot?	No	✗
	Late	 	Is the parrot behind Po?	No	✗

 Boots  Po

Methods

Participants. We recruited seventy 3-year-old children (female = 32, mean age = 40.7 months, range = 36-47 months) and fifty older children (female = 25, mean age = 62.6 months, range = 54-71 months). We planned to recruit only 5-year-old children in the older group but, due to recruitment difficulties, 11 children were younger than 5. To avoid confusion, we thus refer to the two groups as younger and older children, respectively. All were native English speakers, although it was reported that 8 children also heard a second language at home. Twelve younger and two older children were excluded because they had a speech delay (N=1), didn't understand how to play the game (N= 1), did not pay attention or showed no interest in the game (N= 10), or due to experimenter error (N= 2). We also tested forty-eight English-speaking adults from the University of Edinburgh (female = 30, mean age = 20.6 years, range = 18-35 years).

Sample sizes were based on unpublished work from our laboratory, which showed that lexical factors affect the timing of turn taking in a sample of 24 adults. That paradigm was within-subjects, while the present paradigm was between-subjects, and so we doubled the sample size. The sample sizes for children are larger to counter the fact that children did not always complete the full experiment. Data collection stopped when the first author (the experimenter) reached the end of her degree. At that point, we had collected roughly as many child trials (across the two age groups) as adult trials. All adult participants and child caregivers gave written consent to taking part, and children provided verbal assent. The procedure was approved by the Ethics Committee of the School of Psychology, Philosophy, and Language Sciences, University of Edinburgh.

Materials and Procedure. The iPad game was written using Swift 1.2. Participants were told the aim of the game was to help Peter Pan navigate a set of mazes while searching for

either one or two animals, and that they should answer his questions as quickly as possible. Participants completed four mazes (i.e., blocks of trials), and each maze contained 24 forks (i.e. trials).

The number of animals (Maze Type) was varied between participants, and the identity of the animal used in one-animal mazes (parrot or tiger) was counterbalanced across participants. Information Structure (Early: *Is Po hiding the parrot* vs. Late: *Is the parrot behind Po?*) was manipulated within participants.

Between trials, participants moved Peter Pan around the maze using the touchscreen. A trial began when they reached a fork in the maze, whereupon the game froze and a zoomed-in version of the fork appeared, with a cartoon character marking each direction. In two-animal mazes, the two animals Peter Pan was looking for were each hiding behind one of the cartoon characters. In one-animal mazes, the animal that Peter Pan was looking for was hiding behind one of the cartoon characters, while the other cartoon character hid nothing (see Figure 1). As soon as the zoomed-in version of the maze appeared, Peter Pan solicited the participant's help (*Can you help me?*) and after 3.5 seconds he asked the critical question (e.g., *Is the parrot behind Po?*). On half of the trials Peter Pan's question required an affirmative answer, and on half the trials it required a negative answer (e.g., *Is Boots hiding the parrot?* or *Is the parrot behind Boots?* in relation to Figure 1). We chose simple yes/no answers to minimize the complexity of the child's response, on the assumption that this would maximise power to detect concurrent prediction and preparation.

The experimental trials used forty-eight well-known cartoon characters (12 per maze). Since participants tend to answer earlier when questions are longer (Corps, et al., 2018; De Ruiter, Mitterer, & Enfield, 2006), we controlled for question length (both the actual length and the length potentially predicted by the participant). On half the trials, both characters had

short names (1 or 2 syllables; mean length=608ms) and on the other half both had long names (3+ syllables; mean length=1037ms), and this was true both for early and late questions.

Parrot and *tiger* were chosen as animal names because they are length-matched and have similar log frequencies (4.75 and 4.62, respectively) in the children's television subsection of the SUBTLEX-UK database (Van Heuven, Mandera, Keuleers, & Brysbaert, 2014). Finally, overall differences in length between early and late questions were controlled for by the experimental design, as we compared across one- and two-animal mazes.

A female speaker of Scottish-English recorded the 192 possible questions (1 per character per condition) in a quiet room, using slow, child-directed prosody. After recording, a 300ms pause was inserted before the last word. This was done to ensure that participants had sufficient time available to predict the final word and use that prediction to prepare a response. The pause did not sound unnatural, and was in keeping with the slowness of the child-directed speech used in the recordings. However, while the inclusion of the pause should have reduced performance demands on participants, it does also partially limit the generalizability of our findings to more naturalistic conversations.

Participants completed two practice trials before the study began. Before each block, they were shown each of the cartoon characters that they would see in that block; the characters were named by the experimenter and repeated out loud by the participants. The block only began once the experimenter judged that the participant knew each character's name. Participants' spoken responses were recorded using the internal iPad microphone and were coded off-line. Regrettably, copies of the visual stimuli are only available upon request, because the cartoon characters used are restricted by copyright, but the audio stimuli can be found at <https://osf.io/kcp9z/>.

Coding. Audio recordings were coded using Praat by the first and second author, and three trained research assistants. Response times were measured from question offset to the onset of the first speech sound. If the answer was preceded by a filled pause, we measured to the onset of the filled pause to ensure greater consistency (as the filled pause was often coarticulated with the answer); pre-speech in-breaths were excluded, as these were not reliably picked up by the microphone. Responses were also transcribed to determine accuracy. Trials were excluded if participants failed to answer the question, if the trial's question had to be repeated, or if there was too much background noise to allow coding. In total, 14.3% of trials were excluded (Adults: 2.2%, older children: 15.3%, younger children: 25.5%).

Data analysis. Conventionally, analyses of chronometric experiments evaluate whether different experimental conditions cause a difference in the mean response time, and assume that the distribution of response times across conditions has an identical shape. However, as Figure 2A makes clear, the shape of the distribution of response times can be markedly different between conditions. In this experiment for example, the distribution of child response times was much more right-skewed than the distribution of adult response times. When using standard statistical analyses, these large differences in the distributions could, potentially, mask differences in mean response times or indeed induce spurious mean differences. Thus, rather than simply compare condition means, we analysed how the distribution of response times varied across conditions.

In order to capture the shape of both the adult and the child distributions, we assumed that the response times followed an ex-Gaussian distribution (Balota & Yap, 2011), which is the convolution of a Gaussian and an exponential distribution (and thus subsumes the Gaussian as a special case). The ex-Gaussian distribution is known to provide an excellent fit to empirical response time distributions, particularly because it can model their long right tail.

The distribution has three parameters: μ , the location of the Gaussian component; σ , the spread (standard deviation) of the Gaussian component; and τ , which is the inverse rate of the exponential component and which accounts for the thickness of the right tail. Importantly, these three parameters can be differentially affected by one and the same experimental manipulation (Balota & Yap, 2011), which reflects the fact that response times are the result of a combination of multiple processing steps as well as more unusual events (e.g., distractions, cognitive overload), which can cause highly delayed responding. For example, an effect on τ in the absence of an effect of μ could indicate that an experimental manipulation does not delay typical processing steps but does increase the degree of distraction. A distributional analysis should thus provide a much more complete picture of how prediction and formulation affect response times during turn-taking.

Distributional analyses typically require many observations per subject to produce robust subject-level estimates (at least 50 per condition), more than could feasibly be provided by a three-year-old. We overcame this limitation by building on recent advances in Bayesian statistics and distributional multilevel regression modelling (Bürkner, 2016; Carpenter et al., 2016), which allowed us to fit an ex-Gaussian model to the data in a multilevel fashion, using individual observations across all subjects to leverage our by-subject estimates (Gelman A. & Hill, 2007). In particular, we tested how the parameters of the model – the Gaussian's location μ and spread σ , and the exponential rate τ – varied across the different conditions, while accounting for how the observations were hierarchically clustered within subjects. This analysis is analogous to a linear mixed-effects regression, which models how the location (mean) of a distribution varies across conditions and subjects, except that this analysis simultaneously models how all three parameters of the ex-Gaussian distribution vary across those predictors.

We used the same set of predictors to model each of the three parameters: fixed effects of Maze Type (i.e., One vs. Two animals), of Information Structure (Early vs. Late), and of age, along with their full set of interactions; we also included control predictors for the length of the question's final word, and for whether participants had to respond affirmatively or negatively. Factorial predictors were contrast coded (-0.5, 0.5) and continuous predictors were standardized. The random effects structure included random intercepts for each subject and random slopes for Information Structure (which was the only within-subjects factor); the correlation between random slopes and intercepts was fixed to zero. To aid model convergence, we pooled the data for younger and older children (thus, age was a two-level factor, Child vs. Adult), but also explored potential developmental changes in follow-up analyses (see online Supplementary material).

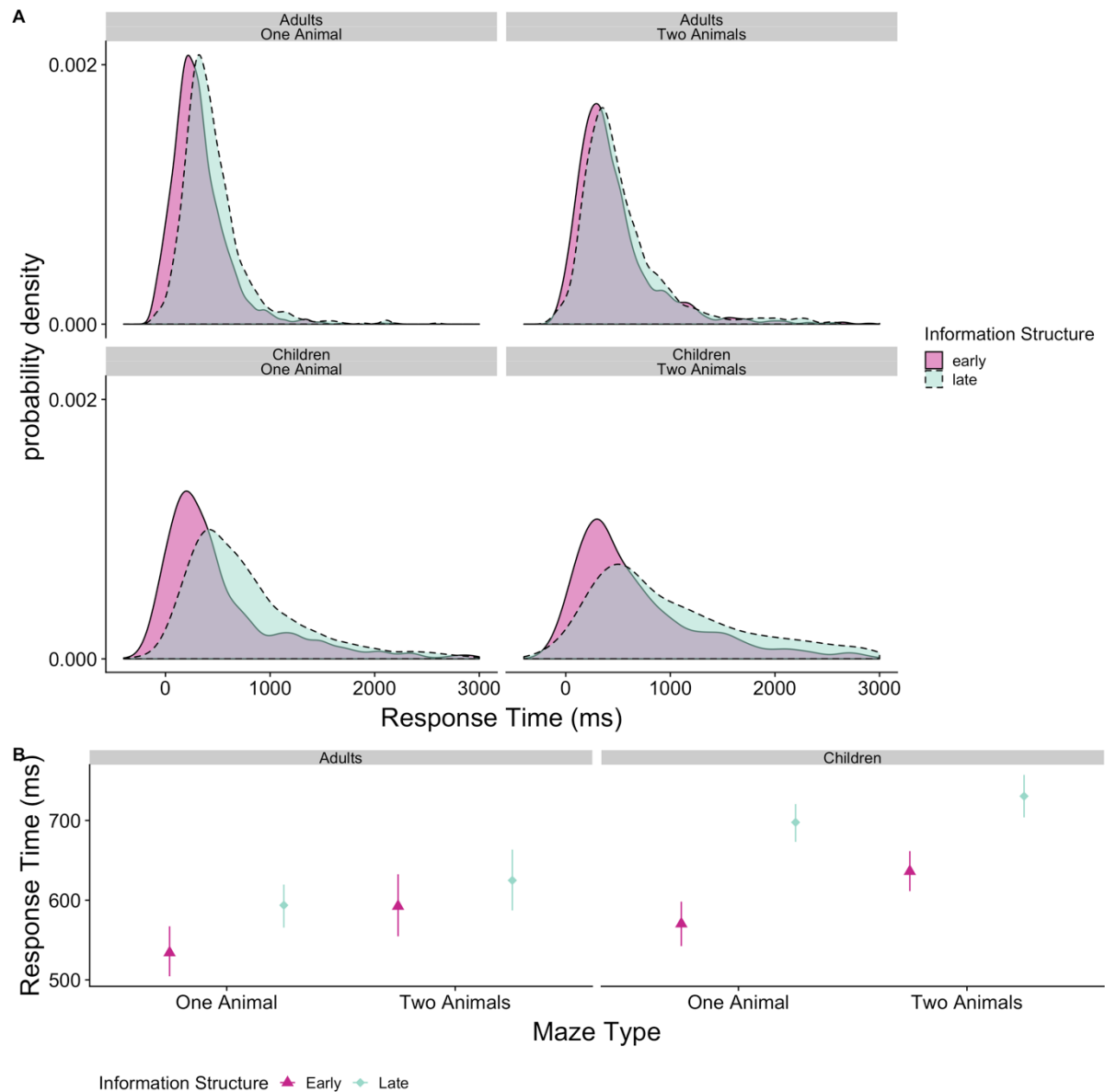
Ex-Gaussian analyses were run using the *brms* package, version 2.1.0 (Bürkner, 2016). We ran 4 chains per model, each for 2000 iterations, with a warm-up period of 1000 iterations, and initial parameter values set to 0. The model converged with no divergent transitions (all \hat{R} values ≤ 1.01). For each parameter, we report estimates (B), estimated error (EE), and the 95% credible interval (CrI). If zero lies outside the credible interval, then we conclude there is sufficient evidence to suggest the estimate is different from zero. Note that σ and τ were fit on the log scale.

Given the novelty of this procedure, we also report a more standard analysis for comparison, using a linear mixed effects model with a Gaussian link function to predict log-transformed response times (fit using the *lmer* function of the *lme4* package, version 1.1-13; Bates, Maechler, Bolker, & Walker, 2015). This had the same fixed effect structure described above, but also included random intercepts for items, and by-item random slopes for Information Structure and Maze Type. We report coefficient estimates (B), standard errors

(SE), and t values for each predictor; 95% CI (Confidence Intervals) are from the *confint* function (method="Wald").

Prior to ex-Gaussian analysis, we excluded trials on which participants provided incorrect answers (Adults: 0.8%, older children: 4.2%, younger children: 17.9%), and trials that were outliers on the left hand side of the distribution, i.e., very early anticipatory responses that were further than 1.5 standard deviations below the age-appropriate mean (<1% of data points at each age group; this excluded roughly the same amount of data as applying a 2.5 standard deviation cut-off to unskewed data). Five younger children were then excluded because they did not provide at least 20 data points after exclusions; no participants were excluded from the other two age groups. Disfluent but correct answers were included in the analyses to reduce data loss. RTs were then standardised and a constant was added to all data points to avoid negative values, since there was a small (<1%) percentage of overlaps. In addition, for the linear model analyses we excluded all RTs longer than 1500ms, in order to further reduce the skew of the distribution. All analyses were conducted in RStudio (version 1.0.143). Data and analysis scripts are available at <https://osf.io/kcp9z/>.

Figure 2. (A) Distribution of response times by Age Group, Information Structure and Maze Type. The Adult panels combined are based on 4528 data points, while the Child panels combined are based on 4863 data points (B) Mean response times (after excluding data points $> 1500\text{ms}$, as in the Gaussian analyses) by age and condition. Error bars represent 95% by-participant CIs bootstrapped over 1000 samples.



Results

If participants prepare a response as they listen to a question, then we expect an interaction between Maze Type and Information Structure, such that response times shift closer to zero when questions mention critical information early rather than late *and* the maze is predictable (one-animal maze) rather than unpredictable (two-animal maze). We thus tested for the presence of this interaction in both adults and children. For the distributional analysis, we assumed that this interaction would shift the location of the Gaussian component (an effect on the μ parameter), but also analysed whether it would affect the spread of the Gaussian (i.e., its standard deviation, or the σ parameter) as well as the rate of the exponential component (the τ parameter), capturing the thickness of the right tail of the distribution.

Figure 2, panel A, shows how the distribution of response times varied across the different age groups and conditions, while Figure 2, panel B, shows the variation in mean response times (see Figure S1 in the Online Supplementary Material for a breakdown of the child data into younger and older children). Our analysis first focused on the location of the Gaussian component of these distributions (μ , Table 1). Overall, this was shifted earlier in time when the maze was predictable (Maze Type: $B=0.027$, $EE=0.013$, $CrI=[0.003,0.053]$), and also when the critical cartoon character's name was mentioned earlier (i.e., when the sentence-final word was *tiger* or *parrot*; Information Structure: $B=0.068$, $EE=0.004$, $CrI=[0.059,0.077]$). This latter main effect was unexpected; we suggest it occurred because participants were faster to recognise the final word when it named one of the animals, since those labels were repeated on every trial of the study.

Crucially, however, these findings were qualified by the predicted interaction, such that information structure had a significantly larger effect on response times when the maze was predictable (one-animal) than when the maze was unpredictable (two-animal; $B=-0.026$,

EE=0.009, CrI=[-0.043,-0.008]). Importantly, there was no further interaction between Information Structure, Maze Type, and Age ($B=0.005$, $EE=0.018$, $CrI=[-0.031,0.039]$), i.e., both children and adults appeared to prepare a response ahead of time based on their predictions about the content of a question. This interaction is perhaps best appreciated by examining Figure 2A and comparing the distributions of responses to early questions (solid line) and late questions (dashed line) in each of the left-hand panels (for predictable one-animal mazes), as compared to the relative distributions of these responses in the corresponding right-hand panel (which show unpredictable two-animal mazes). Critically, for both adults and children, the response distributions from predictable one-animal mazes show considerably less overlap than the response distributions from unpredictable two-animal mazes, indicating how prediction allowed children to formulate a response in advance, and thus reduce response time. Follow-up analyses focused only on the child sample confirmed this result: children as a whole showed the same interaction, and the size of the interaction did not vary across age groups, providing no strong evidence for major developmental differences in the ability to coordinate prediction and production (see Tables S1 and S2 in the Online Supplementary Material). Thus, our findings support a theory of turn-taking development in which the ability to use prediction to flexibly time formulation of a response is an early milestone.

Table 1. Parameters from the Ex-Gaussian distributional analysis, comparing children to adults.

Predictor	μ			σ			τ		
	B(EE) ^a	CrI ^b	\hat{R}^c	B(EE) ^a	CrI ^b	\hat{R}^c	B(EE) ^a	CrI ^b	\hat{R}^c
Intercept	0.477(0.006)	0.465, 0.489	1.003	-3.075(0.042)	-3.158,-2.993	1.000	-1.676(0.046)	-1.769,-1.588	1.001
Age	0.027(0.012)	0.002, 0.051	1.007	-0.222(0.086)	-0.387,-0.048	1.001	-1.187(0.096)	-1.374,-1.001	1.003
Information Structure (IS)	0.068(0.004)	0.060, 0.077	1.000	0.298(0.057)	0.190, 0.411	1.000	0.151(0.036)	0.080, 0.224	0.999
Maze Type (MT)	0.027(0.013)	0.003, 0.053	1.000	0.106(0.083)	-0.057, 0.272	1.000	0.258(0.095)	0.072, 0.453	1.006
Final Word Len	-0.023(0.001)	-0.025,-0.020	1.000	0.107(0.022)	0.063, 0.152	1.000	-0.097(0.013)	-0.124,-0.071	1.000
Answer Type	0.016(0.002)	0.012, 0.021	1.000	-0.016(0.048)	-0.111, 0.078	1.000	0.048(0.026)	-0.003, 0.099	1.000
Age:IS	-0.064(0.009)	-0.081,-0.046	1.001	-0.193(0.109)	-0.410, 0.027	1.000	0.000(0.073)	-0.139, 0.143	1.001
Age:MT	-0.009(0.025)	-0.058, 0.039	1.002	-0.408(0.164)	-0.736,-0.097	1.000	0.009(0.191)	-0.359, 0.395	1.005
IS:MT	-0.026(0.009)	-0.043,-0.008	1.000	0.036(0.112)	-0.186, 0.256	1.000	0.081(0.074)	-0.065, 0.227	1.000
Age:IS:MT	0.005(0.018)	-0.031, 0.039	1.000	-0.361(0.219)	-0.794, 0.075	0.999	-0.151(0.146)	-0.436, 0.137	1.000

^a Estimate (Estimated Error).

^b Credible Interval.

^c A measure of convergence of the algorithm ($\hat{R} = 1$ at convergence).

These key findings were replicated in the (non-distributional) linear mixed-effects analysis. When adults were compared to children (Table 2), the model revealed an interaction between Information Structure and Maze Type ($B = -0.064$, $SE = 0.020$, $t = -3.13$; $CI = [-0.104, -0.024]$), but no further interaction with Age ($B = 0.031$, $SE = 0.030$, $t = 1.03$; $CI = [-0.028, 0.091]$), and the same was true in the analyses that compared younger to older children (see Tables S3 and S4 in the Supplementary Online Material). Thus, both distributional and traditional analyses confirm that, during conversation, children can prepare a response ahead of time, based on a prediction about what their partner will likely say next.

Table 2. Parameters from the Gaussian linear-mixed effect analysis, comparing children to adults.

Predictor	B	SE	t value	lowerCI	upperCI
(Intercept)	-0.493	0.013	-37.038	-0.519	-0.467
Information Structure	0.127	0.014	8.917	0.099	0.155
Maze Type	0.079	0.025	3.196	0.031	0.127
Age	-0.110	0.024	-4.526	-0.158	-0.063
Final Word Len	-0.062	0.006	-11.142	-0.073	-0.051
Answer Type	0.045	0.013	3.443	0.019	0.070
Information Structure:Maze Type	-0.064	0.020	-3.131	-0.104	-0.024
Information Structure:Age	-0.095	0.015	-6.177	-0.125	-0.065
Maze Type:Age	-0.007	0.049	-0.152	-0.103	0.088
Information Structure:Maze Type:Age	0.031	0.030	1.025	-0.028	0.091

Returning to the distributional analysis, we assessed how the different experimental manipulations affected the spread of the Gaussian component, as well as the rate of the exponential component (i.e., right skew), and whether those effects differed between adults and children. As expected based on both observational studies (Casillas et al., 2016) and inspection of Figure 2A, both the spread and skew parameter were larger in children than in adults, indicating that children showed

more variation around the modal response time ($B=-0.222$, $EE=0.086$, $CrI=[-0.387,-0.048]$) as well as a considerably thicker right tail (i.e., highly delayed responses, $B=-1.187$, $EE=0.096$, $CrI=[-1.374,-1.001]$).

In addition, across age groups, the spread and skew were also larger when the critical information came late (effect of Information Structure: spread, $B=0.298$, $EE=0.057$, $CrI=[0.190,0.411]$; skew, $B=-0.035$, $EE=0.036$, $CrI=[0.080,0.224]$). When the maze was not predictable (two animals, right panels of Figure 2A), response times showed larger skew (effect on the rate of the exponential, $B=0.258$, $EE=0.095$, $CrI=[0.072,0.453]$) and they also showed a larger spread, but only in children (interaction between Age and Maze Type, $B=-0.408$, $EE=0.164$, $CrI=[-0.736,-0.097]$). However, and importantly, neither parameter showed a reliable interaction between Information Structure and Maze Type (spread, $B=0.036$, $EE=0.112$, $CrI=[-0.186, 0.256]$; skew, $B=0.081$, $EE=0.074$, $CrI=[-0.065,0.227]$). This suggests that flexible predictive formulation of responses mainly acts to shift the location of the response time distribution, while it does not affect response variability or the likelihood of very delayed responding.

A notable feature of these latter analyses is that they highlight how attention to the distribution of response times can be more informative than attention to their mean. Our traditional analysis of the mean suggested that children's turn-taking was significantly slower than adults' (a significant Age-related decrease; see Figure 2b and Table 2), but the distributional analysis showed that this was not simply because adults' processing speed is greater. Rather, children and adults had similar modal response times (see Figure 2a and Table 1)¹, but children also showed significantly more variability in

¹ If anything, the model in Table 1 suggests modal response times (μ) are actually *greater* in adults, but this likely results from a trade-off with the concurrent, much larger decrease in τ .

their response times, and were also more prone to respond after very long delays (main effects of age on both σ and τ parameters, see Figure 2a and Table 1). This pattern suggests that the major developmental change in turn-taking is not in the speed with which children coordinate comprehension and formulation, but in how likely children are to become distracted, or experience difficulties in switching between tasks. We return to this issue in the Discussion.

Finally, we assessed how response times were affected by our two control predictors: the length of the question's final word, and whether the participant had to answer Yes or No. Consistent with prior work by De Ruiter et al (2006), when the final word was longer, then not only were response times shifted closer to zero, but the skew of those responses was reduced (effects on the location, $B=-0.023$, $EE=0.001$, $CrI=[-0.025,-0.020]$, and the rate of the exponential, $B=-0.097$, $EE=0.013$, $CrI=[-0.124,-0.071]$). This finding is important because it confirms that children do not typically wait until the question is over to begin formulation; rather, they start formulation as they listen to the question, and thus respond earlier when the longer final word gives them more time to prepare. Interestingly, when the final word was longer, the spread of the Gaussian component tended to be somewhat larger ($B=0.107$, $EE=0.022$, $CrI=[0.063,0.152]$). This is consistent with other evidence showing that adults respond earlier to questions that are longer, but that the timings of their answers are less precise, e.g., resulting in more overlaps (Corps et al., 2018). Finally, when participants responded No, their response times were shifted later compared to when they responded Yes (effect on the location parameter, $B=0.016$, $EE=0.002$, $CrI=[0.012,0.021]$), but there were no further effects on the spread or skew of the distribution. This latter finding is consistent with some corpus evidence (Stivers et al., 2009) and could be explained by the fact that rejections are dis-preferred responses (Kendrick & Torreira, 2015).

Discussion

Observational analyses of preschool children's conversational turn-taking suggest that it is poorly timed, but our experiment reveals that preschoolers' ill-timed responses mask a sophisticated ability to coordinate language comprehension and language production: They can generate predictions about what their conversational partners will say, use those predictions to prepare a response while still listening, and thus respond more quickly. In this way, children act to optimize the timing of their conversational turns.

Our findings contribute to a growing body of work highlighting the importance of turn-taking for human culture and development. Most notably, they are consistent with the ethological approach to turn-taking advocated by Levinson (2006; 2016), which argues for continuity in the mechanisms that allow young infants to engage in well-timed non-linguistic turn-taking, and the mechanisms that preschoolers use when engaging in (less well-timed) linguistic turn-taking (Hilbrink et al., 2015). Such continuity acts as a key argument in support of Levinson's stronger claims: that turn-taking represents a universal biological adaptation in humans (rather than, e.g., a cultural adaptation for facilitating smooth linguistic interaction); and that, over historical time, languages have evolved to ensure that conversational turn-taking is as smooth as possible (e.g., by favouring grammatical or prosodic features that allow listeners to quickly identify if a turn is a question or statement).

The present data also shed light on the striking recent discovery that children's longitudinal language development is better predicted by the quality of their early conversational experiences than by more traditional measures, such as quantity of words heard (Hirsh-Pasek et al., 2015; Zimmerman et al., 2009). The particular contribution of our study is to specify one of the cognitive mechanisms, the flexible coordination of prediction and formulation, that would allow children to engage their caregivers in these higher-quality conversational interactions. For instance, this mechanism allows children to generate conversational turns that both build informatively on the previous speaker (e.g., by responding to a

request), and that onset within a pragmatically appropriate timeframe. Thus, children can actively promote caregiver engagement and maximise their own opportunities to learn. In turn, the timeliness of a child's responses allows adults to monitor their degree of understanding, and thus provide appropriate feedback. In combination, these factors should act as scaffolding for children as they acquire new linguistic and world knowledge from conversations (see Clark, 2007; Gelman, S. 2009), and may also help foster a better understanding of the social uses of language (Dunn & Shatz, 1989).

One question raised by our study is why, if children can coordinate comprehension and production, they still often produce poorly-timed conversational turns. Levinson (2016) suggested that children's poor timing may reflect difficulties that they have in producing complex language (consistent with naturalistic observations; Casillas et al., 2016; Clark & Lindsey, 2015), which is partially consistent with the present data: Here, children only needed to produce very simple Yes/No answers, and their modal response times were actually strikingly similar to adults'. However, to fully understand how production difficulties constrain the development of conversational timing, it will be important to test more complex responses, as well as younger children. Moreover, even when children in our task produced simple utterances, they still often left long gaps (i.e., children's response distributions had a larger right tail than adults'), which suggests that factors other than linguistic planning difficulty may also contribute to the poor timing of children's linguistic turn-taking.

We thus propose that such poor timing may also arise from a more general factor, such as distractibility or skill at switching tasks (e.g., from comprehension to formulation, Zelazo et al., 2003). This would be consistent with previously reported associations between executive processes and the weight of the right tail of the response distribution (Schmiedek, Oberauer, Wilhelm, Süß, & Wittmann, 2007, though see Matzke & Wagenmakers, 2009). It also highlights an important message from the present paper: Distributional data analysis methods (Balota & Yap, 2011; Umlauf & Kneib, 2018), which

characterise the shape of the response time curve, allowed us to capture children's coordination skills while at the same time accounting for (and indeed modelling) the noise in their response times. This underscores the utility of these analytic methods for the interpretation of chronometric experiments.

Conclusion

Smooth conversational turn-taking appears to be universal, despite the pressure that it places on understanding and using language (Levinson, 2016). Our findings suggest that preschoolers, who are still learning to use complex language, can nonetheless flexibly coordinate language comprehension and language production, and thus optimize the timing of their conversational interactions. The early development of these skills suggests that an understanding of turn-taking will shed light on language and cognitive development more broadly.

Acknowledgements

This research was supported by the Carnegie Foundation and by ESRC Studentship N71737J to LL, by the Leverhulme Trust (RPG 2014-253) to HR and Martin Pickering, and by an ESRC Future Research Leaders award (ES/L01064X/1) to HR.

References

- Balota, D. A., & Yap, M. J. (2011). Moving beyond the mean in studies of mental chronometry: The power of response time distributional analyses. *Current Directions in Psychological Science*, 20(3), 160-166.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1-48.

- Bögels, S., Magyari, L., & Levinson, S. C. (2015). Neural signatures of response planning occur midway through an incoming question in conversation. *Scientific Reports*, doi:10.1038/srep12881
- 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.
- Bürkner, P.-C. (2016). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, 80(1), 1-28.
- Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., . . . Riddell, A. (2016). Stan: A probabilistic programming language. *Journal of Statistical Software*, 20(2), 1-37.
- Casillas, M., Bobb, S. C., & Clark, E. V. (2016). Turn taking, timing, and planning in early language acquisition. *Journal of Child Language*, 43(6), 1310-1337.
- Casillas, M., & Frank, M. C. (2017). The development of children's ability to track and predict turn structure in conversation. *Journal of Memory and Language*, 92, 234-253.
- Clark, E. V. (2007). Young children's uptake of new words in conversation. *Language in Society*, 36(2), 157-182.
- Clark, E. V., & Lindsey, K. L. (2015). Turn-taking: A case study of early gesture and word use in answering WHERE and WHICH questions. *Frontiers in Psychology*, doi:10.3389/fpsyg.2015.00890
- Corps, R., Crossley, A., Gambi, C., & Pickering, M. J. (2018). Early preparation during turn-taking: listeners use content predictions to determine what to say but not when to say it. *Cognition*, 175, 77-95.
- De Ruiter, J. P., Mitterer, H., & Enfield, N. J. (2006). Projecting the end of a speaker's turn: A cognitive cornerstone of conversation. *Language*, 82(3), 515-535.

- Dunn, J., & Shatz, M. (1989). Becoming a conversationalist despite (or because of) having an older sibling. *Child Development*, 60(2), 399-410.
- Gambi, C., Pickering, M. J., & Rabagliati, H. (2016). Beyond Associations: Sensitivity to structure in pre-schoolers' linguistic predictions. *Cognition*, 157, 340-351.
- Garvey, C., & Berninger, G. (1981). Timing and turn taking in children's conversations 1. *Discourse Processes*, 4(1), 27-57.
- Gelman A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge, UK: Cambridge University Press.
- Gelman, S. A. (2009). Learning from others: Children's construction of concepts. *Annual Review of Psychology*, 60, 115-140.
- Hilbrink, E. E., Gattis, M., & Levinson, S. C. (2015). Early developmental changes in the timing of turn-taking: a longitudinal study of mother–infant interaction. *Frontiers in Psychology*, doi:10.3389/fpsyg.2015.01492
- Hirsh-Pasek, K., Adamson, L. B., Bakeman, R., Owen, M. T., Golinkoff, R. M., Pace, A., . . . Suma, K. (2015). The contribution of early communication quality to low-income children's language success. *Psychological Science*, 26(7), 1071-1083.
- Indefrey, P., & Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition*, 92(1), 101-144.
- Jaffe, J., Beebe, B., Feldstein, S., Crown, C. L., & Jasnow, M. D. (2001). Rhythms of dialogue in infancy: Coordinated timing in development - Introduction. *Monographs of the Society for Research in Child Development*, 66(2), i-viii + 1-149.
- Keitel, A., Prinz, W., Friederici, A. D., von Hofsten, C., & Daum, M. M. (2013). Perception of conversations: The importance of semantics and intonation in children's development. *Journal of Experimental Child Psychology*, 116(2), 264-277.

- Kendrick, K. H., & Torreira, F. (2015). The timing and construction of preference: A quantitative study. *Discourse Processes*, 52(4), 255-289.
- Levinson, S. C. (2006). On the human interaction engine. In N. J. Enfield & S. C. Levinson (Eds.), *Roots of Human Sociality* (pp. 39-69). Oxford: Berg.
- Levinson, S. C. (2016). Turn-taking in human communication—origins and implications for language processing. *Trends in Cognitive Sciences*, 20(1), 6-14.
- Magyari, L., Bastiaansen, M. C., de Ruiter, J. P., & Levinson, S. C. (2014). Early anticipation lies behind the speed of response in conversation. *Journal of Cognitive Neuroscience*, 26(11), 2530-2539.
- Mani, N., & Huettig, F. (2012). Prediction during language processing is a piece of cake - but only for skilled producers. *Journal of Experimental Psychology: Human Perception and Performance*, 38(4), 843-847.
- Matzke, D., & Wagenmakers, E.-J. (2009). Psychological interpretation of the ex-Gaussian and shifted Wald parameters: A diffusion model analysis. *Psychonomic Bulletin & Review*, 16(5), 798-817.
- Murray, L., & Trevarthen, C. (1986). The infant's role in mother–infant communications. *Journal of Child Language*, 13(1), 15-29.
- Rabagliati, H., Gambi, C., & Pickering, M. J. (2015). Learning to predict or predicting to learn? *Language, Cognition, and Neuroscience*, 31(1), 94-105.
- Riest, C., Jorschick, A. B., & de Ruiter, J. P. (2015). Anticipation in turn-taking: mechanisms and information sources. *Frontiers in psychology*, doi:10.3389/fpsyg.2015.00089
- Sacks, H., Schegloff, E.A., & Jefferson, G. (1974). A simplest systematics for the organization of turn taking in conversation. *Language*, 50(4), 696-735.

- Schmiedek, F., Oberauer, K., Wilhelm, O., Süß, H.-M., & Wittmann, W. W. (2007). Individual differences in components of reaction time distributions and their relations to working memory and intelligence. *Journal of Experimental Psychology: General*, 136(3), 414-429.
- Sjerps, M. J., & Meyer, A. S. (2015). Variation in dual-task performance reveals late initiation of speech planning in turn-taking. *Cognition*, 136, 304-324.
- Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., . . . Levinson, S. C. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences*, 106(26), 10587–10592.
- Stivers, T., Sidnell, J., & Bergen, C. (2018). Children's responses to questions in peer interaction: A window into the ontogenesis of interactional competence. *Journal of Pragmatics*, 124, 14-30.
- Umlauf, N., & Kneib, T. (2018). A primer on Bayesian distributional regression. *Statistical Modelling*, 18(3-4), 219-247.
- Van Heuven, W. J., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology*, 67(6), 1176-1190.
- Zelazo, P. D., Müller, U., Frye, D., Marcovitch, S., Argitis, G., Boseovski, J., . . . Sutherland, A. (2003). The development of executive function in early childhood. *Monographs of the Society for Research in Child Development*, 68(3), 1-151.
- Zimmerman, F. J., Gilkerson, J., Richards, J. A., Christakis, D. A., Xu, D., Gray, S., & Yapanel, U. (2009). Teaching by listening: The importance of adult-child conversations to language development. *Pediatrics*, 124(1), 342-349.

Supplementary Analyses

The analyses reported in the main text pooled data from younger and older children, a decision that we made because the total number of data points from each age group only amounted to about half the data points that we collected from adults (due to various factors, such as children not completing the experiment, or providing incorrect answers, see Figure S1 below). When the children's data were pooled the resulting estimates were more robust (and convergence issues were removed).

However, we also wanted to check for any developmental trends within the child group, as well as to confirm that, when children's data alone were analyzed, the critical interaction between maze type and information structure was still present. We thus conducted two follow-up analyses. The first follow-up analysis compared younger to older children (i.e., using age as a categorical predictor), while the other included the child's age in months as a (centred and scaled) continuous predictor.

The results of these distributional follow-up analyses on the child data, using the ex-Gaussian distribution, can be found in Tables S1 (categorical) and S2 (continuous) below. They fully confirm the distributional analyses reported in the main text. Critically, we again found that the distribution of response times across conditions was shifted in line with an interaction between Information Structure and Maze Type ($B = -0.035$, $EE = 0.013$, $CrI = [-0.062, -0.009]$), again without any further modulation by Age ($B = -0.018$, $EE = 0.026$, $CrI = [-0.069, 0.034]$; see Table S1), showing that both younger and older children can predictively prepare a response to a question while conversing (compare the middle to the bottom panels in Figure S1A, or the middle and the right-hand panels in Figure S1B below). These results also held when age was analyzed as a continuous variable (see Table S2).

Similarly, these key findings were replicated in follow-up (non-distributional) linear mixed-effects analysis of the child data, which are reported in Tables S3 (age as a categorical predictor) and S4 (age as a continuous predictor). The model that compared older to younger children showed a significant interaction between Information Structure and Maze Type ($B = -0.061$, $SE = 0.028$, $t = -2.14$; $CI = [-0.117, -0.005]$), but no further interaction with Age ($B = -0.039$, $SE = 0.046$, $t = -0.84$; $CI = [-0.129, 0.052]$; see Table S3), and the same was true for the model that treated age as continuous (see Table S4).

In sum, follow-up analyses found no strong evidence for a developmental trend within the child data, suggesting that even the younger children in our sample possess the ability to coordinate prediction with early formulation of their response, but further work is necessary to confirm this finding with a larger dataset.

Supplementary Tables

Table S1. Parameters from the Ex-Gaussian analysis, comparing younger to older children (Age). Please refer to the caption for Table 1 (main text) for an explanation of column labels.

Predictor	μ			σ			τ		
	B(EE)	CrI	Rhat	B(EE)	CrI	Rhat	B(EE)	CrI	Rhat
Intercept	0.460(0.007)	0.447, 0.474	1.002	-3.088(0.080)	-3.254, -2.938	1.001	-1.070(0.052)	-1.169, -0.968	1.002
Age	0.001(0.014)	-0.026, 0.029	1.005	0.047(0.140)	-0.217, 0.339	1.001	-0.095(0.101)	-0.296, 0.101	1.001
Information Structure (IS)	0.097(0.007)	0.085, 0.111	1.000	0.336(0.109)	0.136, 0.556	1.000	0.162(0.044)	0.076, 0.248	1.000
Maze Type (MT)	0.027(0.014)	-0.001, 0.055	1.008	0.204(0.140)	-0.085, 0.474	1.002	0.261(0.103)	0.060, 0.461	1.011
Final Word Len	-0.043(0.003)	-0.049, -0.037	1.000	0.107(0.048)	0.015, 0.203	1.001	-0.077(0.018)	-0.113, -0.042	1.000
Answer Type	0.029(0.005)	0.019, 0.039	1.000	0.136(0.099)	-0.057, 0.333	0.999	0.006(0.034)	-0.060, 0.076	1.000
Age:IS	-0.012(0.013)	-0.038, 0.015	1.000	-0.145(0.217)	-0.575, 0.263	1.000	-0.084(0.086)	-0.255, 0.082	1.000
Age:MT	-0.015(0.028)	-0.071, 0.040	1.004	0.546(0.276)	0.022, 1.106	1.001	-0.006(0.208)	-0.421, 0.410	1.005
IS:MT	-0.035(0.013)	-0.062, -0.009	1.000	0.329(0.217)	-0.086, 0.754	1.001	0.182(0.087)	0.015, 0.352	1.001
Age:IS:MT	-0.018(0.026)	-0.069, 0.034	1.001	-0.427(0.444)	-1.303, 0.408	1.000	-0.002(0.182)	-0.357, 0.352	1.000

Table S2. Parameters from the Ex-Gaussian analysis using the child's age (in months) as a scaled continuous predictor (contAge). Please refer to the caption to Table 1 (main text) for an explanation of column labels.

Predictor	μ			σ			τ		
	B(E)	CrI	Rhat	B(E)	CrI	Rhat	B(E)	CrI	Rhat
Intercept	0.461(0.007)	0.447, 0.475	1.002	-3.096(0.076)	-3.252,-2.954	1.000	-1.081(0.053)	-1.183,-0.975	1.007
contAge	-0.001(0.007)	-0.015, 0.013	1.003	0.016(0.066)	-0.110, 0.148	1.000	-0.062(0.052)	-0.164, 0.035	1.003
Information Structure (IS)	0.097(0.007)	0.084, 0.111	1.000	0.345(0.106)	0.134, 0.550	1.000	0.152(0.044)	0.068, 0.238	1.001
Maze Type (MT)	0.024(0.014)	-0.005, 0.052	1.006	0.247(0.133)	-0.018, 0.505	0.999	0.263(0.105)	0.058, 0.468	1.000
Final Word Len	-0.043(0.003)	-0.049,-0.037	1.001	0.106(0.046)	0.018, 0.196	1.000	-0.077(0.018)	-0.112,-0.043	1.000
Answer Type	0.029(0.005)	0.019, 0.039	1.000	0.136(0.100)	-0.058, 0.336	1.000	0.006(0.034)	-0.060, 0.074	0.999
contAge:IS	-0.005(0.007)	-0.018, 0.009	1.000	-0.081(0.106)	-0.287, 0.130	1.000	-0.043(0.043)	-0.129, 0.041	1.001
contAge:MT	-0.011(0.014)	-0.039, 0.016	1.000	0.295(0.134)	0.037, 0.563	1.003	-0.037(0.100)	-0.234, 0.162	1.006
IS:MT	-0.036 (0.013)	-0.062,-0.009	1.000	0.325(0.213)	-0.088, 0.741	1.000	0.184 (0.089)	0.006, 0.360	1.000
contAge:IS:MT	-0.010 (0.013)	-0.037, 0.017	1.000	-0.243(0.211)	-0.648, 0.178	0.999	-0.009(0.087)	-0.178, 0.161	0.999

RUNNING HEAD: PRESCHOOLERS' TURN-TAKING

Table S3. Parameters from the Linear Mixed-Effect analysis, comparing younger to older children (Age).

Predictor	B	SE	t value	lowerCI	upperCI
(Intercept)	-0.423	0.016	-26.856	-0.454	-0.392
Information Structure	0.181	0.018	10.119	0.146	0.217
Maze Type	0.089	0.029	3.115	0.033	0.145
Age	-0.019	0.028	-0.667	-0.074	0.036
Final Word Len	-0.075	0.008	-9.880	-0.090	-0.060
Answer Type	0.047	0.015	3.086	0.017	0.077
Information Structure:Maze Type	-0.061	0.028	-2.136	-0.117	-0.005
Information Structure:Age	-0.018	0.023	-0.780	-0.063	0.027
Maze Type:Age	-0.028	0.056	-0.502	-0.139	0.082
Information Structure:Maze Type:Age	-0.039	0.046	-0.836	-0.129	0.052

Table S4. Parameters from the Linear Mixed-Effect analysis, using the child's age (in months) as a scaled continuous predictor (contAge).

Predictor	B	SE	t value	lowerCI	upperCI
(Intercept)	-0.424	0.016	-26.645	-0.456	-0.393
Information Structure	0.196	0.018	11.139	0.161	0.230
Maze Type	0.085	0.029	2.936	0.028	0.142
contAge	-0.016	0.014	-1.141	-0.042	0.011
Final Word Len	-0.075	0.008	-9.903	-0.090	-0.060
Answer Type	0.046	0.015	3.119	0.017	0.075
Information Structure:Maze Type	-0.065	0.029	-2.267	-0.121	-0.009
Information Structure:contAge	-0.005	0.011	-0.472	-0.028	0.017
Maze Type:contAge	-0.028	0.027	-1.028	-0.081	0.025
Information Structure:Maze Type:contAge	-0.019	0.023	-0.830	-0.063	0.026

Supplementary Figure

Figure S1. (A) Distribution of response times by Age Group, Information Structure and Maze Type. The Adult panels combined are based on 4528 data points; the combined number of data points is 2867 for the Older children panels and 1996 for the Younger children panels

(B) Mean response times (after excluding data points $> 1500\text{ms}$, as in the Gaussian analyses)

by age and condition. Error bars represent 95% by-participant CIs bootstrapped over 1000 samples.

