DOI: 10.1111/cdep.12515

ARTICLE

Child Development Perspectives

The role of prediction error in the development of language learning and memory

Jutta Kray[1] | Linda Sommerfeld[1] | Arielle Borovsky[2] | Katja Häuser[1]

[1]Department of Psychology, Saarland University, Saarbrücken, Germany

[2]Department of Speech, Language, and Hearing Sciences, Purdue University, West Lafayette, Indiana, USA

Correspondence
Jutta Kray, Department of Psychology, Saarland University, Campus A 1 3, 66123 Saarbrücken, Germany.
Email: j.kray@mx.uni-saarland.de

Funding information
Deutsche Forschungsgemeinschaft, Grant/Award Number: 232722074–SFB1102

Abstract

Prediction error plays a pivotal role in theories of learning, including theories of language acquisition and use. Researchers have investigated whether and under which conditions children, like adults, use prediction to facilitate language comprehension at different levels of linguistic representation. However, many aspects of the reciprocal relation between prediction error and the development of language learning remain unclear. In this article, we review studies in language development that can inform us about the role of prediction error in updating, learning, and retrieving linguistic information. We argue that the study of individual differences in linguistic and cognitive skills will help the field understand more thoroughly whether, when, and why prediction aids language learning, and whether prediction error necessarily results in language learning and retrieval from memory. We close with a discussion of the needs and challenges for researchers to answer these questions.

KEYWORDS
language development, learning and memory, prediction

Comprehensive accounts of human adaptive behavior have increasingly focused on prediction as a core component that explains individual goals (through predictive coding) and learning (via prediction error). These accounts posit that the human brain generates predictions about future events by continuously updating and integrating bottom-up information from the external and internal environment with top-down expectations that are determined by individual goals or shaped by previous experiences (Clark, 2013). In neuro-cognitive learning theories, predictive coding or prediction error is defined as a basic mechanism that aligns the internal representation of goals with perceived events to guide perception and action (Friston, 2010). The notion of prediction has also influenced theoretical models of human language comprehension, which assume that individuals continuously form expectations about upcoming linguistic contents at the semantic, morpho-syntactic, lexical, and discourse levels. Hence, prediction facilitates language comprehension, which in turn can free up cognitive resources to enable more effortless communication between individuals (Kuperberg & Jaeger, 2016; Pickering & Gambi, 2018).

Theoretical models about prediction are driven primarily by evidence from studies of young adults, who have a comprehensive knowledge base, extensive language experience, and fully developed cognitive systems. In contrast, children have a smaller knowledge base, their language experience changes tremendously over childhood, and their cognitive abilities are increasing substantially. Therefore, developmental studies are an optimal testing environment to answer the questions of whether, when, and under which conditions prediction error serves as a general mechanism to help individuals learn novel word meanings and store semantic representations in memory.

Several excellent reviews address the role of prediction in research on language learning (e.g., Babineau et al., 2023; Zettersten, 2019), including the implementation of prediction errors in computational models of language learning

Abbreviations: EEG, electroencephalogram; MMN, mismatch negativity.

(Rabagliati et al., 2016). Therefore, we will discuss only briefly what we know about predictive language processing in childhood. Our main aim is to highlight an emerging research field that investigates not only how prediction error supports immediate linguistic updating during online processing but also how prediction drives long-term learning and aids in retrieving linguistic representations from memory. We also consider how individual differences in both linguistic skills (e.g., vocabulary size) and cognitive skills (e.g., working-memory capacity) might modulate predictive processing. Cognitive mediators in the use of prediction have been underexplored in the literature, and assumptions about cognitive constraints have not been used in computational models about language development.

PREDICTION-BASED ACCOUNTS IN LEARNING AND LANGUAGE THEORIES

Although prediction-based accounts share the general assumption that expectations reflect previous experiences and knowledge and that these expectations are continuously updated to create optimal cognitive representations of a given situation, the precise notion of prediction differs slightly across research fields. In language research, prediction refers to preactivating and preprocessing incoming spoken or written linguistic information that unfolds in response to the emerging message and critically, in advance of the target linguistic stimulus (Kutas et al., 2014). For example, when listening to the start of a sentence like "*On hot summer days, children love to eat…*," listeners activate the representation of the word *ice cream* and not *broccoli* before hearing the final word. Prediction is probabilistic in nature and can occur at multiple levels of representation. That is, comprehenders may generate expectations for an upcoming linguistic signal at the level of semantic meaning, syntactic structure, or phonological and orthographic information. The preactivation of information is thought to depend not only on the constraints of the context but also on comprehenders' prior knowledge and current goals, as well as on the demands of the current task (for a review, see Kuperberg & Jaeger, 2016).

To a great extent, studies on prediction have been conducted with adults who have a large lexicon and knowledge base. But what about cases where comprehenders' knowledge is emerging, as occurs in early language learning in infancy? Next, we summarize empirical evidence pertaining to whether prediction occurs in infancy and which methods are suitable to investigate prediction.

PREDICTIVE PROCESSING IN EARLY AND MIDDLE CHILDHOOD

Children's brains process the regularity of events and code events that are infrequent or not in line with their expectations. For instance, the mismatch negativity (MMN) is a component of electroencephalograms (EEGs) that is elicited by irregularity in the incoming perceptual input. Children as young as 2 months show the MMN in response to infrequent auditory information, suggesting that they discriminate phonemes to identify word boundaries (Friederici, 2005; Ylinen et al., 2017). Another well-known EEG component, the N400, reflects semantic processing at the word and sentence level, and indicates whether the meaning of a word is known or fits into the sentence context. Even children as young as 9 months show a N400 to words when a presented picture does not match the meaning of a spoken word (Parise & Csibra, 2012). At the sentence level, the N400 has been demonstrated in 19- and 24-month-olds for words that do not fit the context of the preceding sentence (Friedrich & Friederici, 2005). However, coding the irregularity of events (as reflected in the MMN) and measuring integration difficulties (by means of the N400) do not necessarily mean that children generate predictions during online language processing. As noted in comprehensive reviews (e.g., Kuperberg & Jaeger, 2016; Rabagliati et al., 2016; Zettersten, 2019), event-related potential mismatch responses (the MNN or the N400) are consistent with two possible accounts: They reflect prediction-related processing difficulties or they reflect the difficulty in integrating an unpredictable word into a context (see also Kutas et al., 2014). (For the sociodemographic characteristics of the studies reviewed herein, please see Table S1.)

To circumvent this issue, researchers have sought ways to investigate the activation of critical words *before* they become available in the input, for example, by using the visual world paradigm (Huettig & Mani, 2016). In this paradigm, a constraining sentence is presented auditorily in combination with a visual scene consisting of two or more objects. Typically, only one of the depicted objects (i.e., the target) is consistent with a predictable word in the sentence. Predictive processing is assessed by comparing eye movements toward the target versus distractor objects before the target name is realized acoustically (see Figure 1).

Table 1 provides an overview and summary of developmental studies that applied paradigms suitable for measuring predictive processing, sorted by the age range of the participants and levels of linguistic representation. Many of the studies investigated predictive processing at the semantic level of representation. Accordingly, in eye-tracking studies, by age 2, children generated semantic predictions based on sentence context. For instance, children showed a larger increase in target fixation to fit a semantically constraining verb (e.g., *eat*) to a picture conveying the context of a sentence (e.g., *ice cream*) than they did in their ability to fit a semantically neutral verb (e.g., see) to a picture conveying the sentence context, with the increase in target fixations already occurring prior to the final noun (see Figure 1; Mani & Huettig, 2012, 2014). Children, like adults, may even show anticipatory fixations to not just one but three or four visual prediction options (Sommerfeld et al., 2023). Preactivation of semantic features is prioritized in 2-year-olds, although

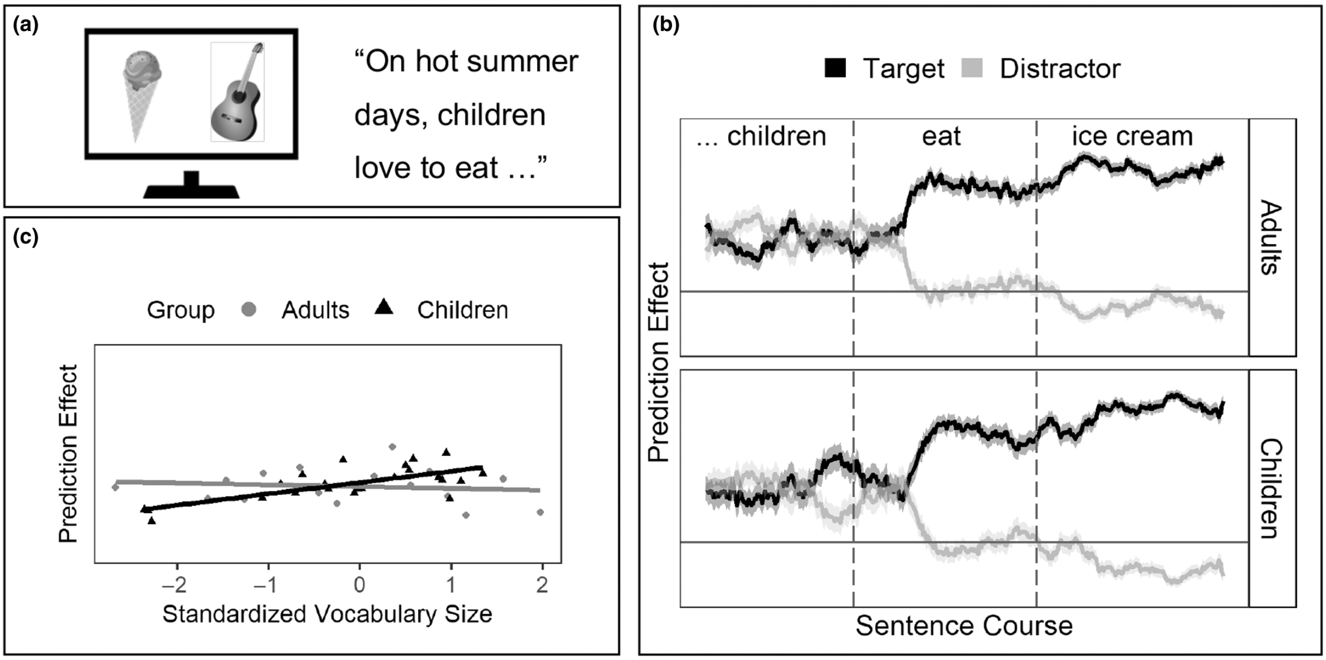

FIGURE 1 Examples of stimuli and results of visual-word eye-tracking studies. (a) Example stimulus of a typical visual-world eye-tracking study on prediction. Participants were auditorily presented with a sentence such as "*On hot summer days, children love to eat ice cream*" while they looked at a visual display presenting one target object (ice cream) and one distractor object (guitar). (b) Example results on the time course of target fixations while hearing the sentence. Upon hearing the verb "*eat*," adults and children anticipatorily fixated on the single *edible* object while fixations on the distractor decreased. This effect occurred shortly after verb presentation and prior to the target presentation, suggesting that children and adults predicted the noun based on the verb. (c) Expected relations between the prediction effect and vocabulary size in groups of children and adults.

they also preactivate other information, such as shape information (Bobb et al., 2016). Between ages 3 and 10, children preactivate lexical-semantic information not just based on a single semantically constraining verb, but also by combining multiple words using syntactic constraints (Borovsky, Elman, & Fernald, 2012). With increasing age, evidence suggests that prediction occurs at the morpho-syntactic (e.g., Cholewa et al., 2019; Günther et al., 2023), syntactic (e.g., Beretti et al., 2020; de Carvalho et al., 2021), and discourse level (e.g., Borovsky, Elman, & Fernald, 2012; Lee et al., 2022; Yazbec et al., 2019).

One major shortcoming of many of these studies is that they examined participants in a restricted age range or one age group, making it difficult to identify developmental changes in predictive processing. That is, studies have focused more on the age at which children predict rather than how this ability changes over time. Studies comparing a group of children with a group of young adults have often found similar patterns of results in both age groups (e.g., Borovsky & Creel, 2014; Gambi et al., 2016; Lee et al., 2022; Lindsay et al., 2019), probably because the complexity of the task and the materials were adapted to children. Hence, while these studies illustrate that even young children have the capacity to predict, we do not know whether children use these skills in the same ways as adults, or how these skills emerge and change over time. Another shortcoming is that many of the studies investigated predictive processing at only one level of representation (e.g., semantic or syntactic), so conclusions are limited regarding the potentially different developmental trajectories using prediction at different levels of representation and how they may interact (cf., Babineau et al., 2023).

In summary, we know that children can use prediction as a unified principle for processing upcoming information at different levels of linguistic presentations (e.g., semantic, lexical, phonological). We also know that the use of prediction depends strongly on a sufficiently detailed representation of prior context. For a comprehensive theoretical framework about the development of prediction and its impact on language processing, we need to learn about the boundary conditions that determine the use of predictions in children. For instance, we know little about how much prior learning is required at which age and at which level of representation. Next, we review new research that applies paradigms that systematically manipulate the exposure to novel linguistic information to either immediately update representations or retrieve them from memory.

RELATIONS BETWEEN PREDICTION ERROR AND SUBSEQUENT LEARNING AND MEMORY

In many neural network models, prediction error is essential for language learning (for a review, see Rabagliati

TABLE 1 Results of studies on predictive processing in early and middle childhood.

Publication	Age group(s)	Level of processing	Methods and measures	Moderators	Results
Mani and Huettig (2012)	• 2 years	Semantical constraints	• Sentences with constraining versus neutral verbs and predictable nouns ("The boy eats/sees the cake") in scenes with target (cake) and distractor (bird) object • Eye tracking: Proportional target fixations (cake) in verb and noun window	• Comprehension vocabulary • Production vocabulary	• 2-year-olds showed more anticipatory target fixations in constraining versus neutral condition. They used semantic cue to predict • Infants with high productive vocabulary showed stronger prediction effect
Bobb et al. (2016)	• 2.5 years	Semantical constraints	• Constraining sentences ("The boy takes his bucket and his …") in scenes with semantic (bucket), shape (fork), and unrelated (hat) primes • Eye tracking: Proportional fixations of object classes in constraining and target window	-	• 2-5-year-olds showed more anticipatory fixations of semantic but also of shape versus unrelated primes • Children used semantic cue to predict semantic and shape information
Gambi, Jindal, et al. (2021)	• 2–5 years	Semantical constraints	• Constraining sentences ("The *dog chews* on the …") in scenes with high (bone), middle (slippers), and low (pajama) predictable objects; additional neutral sentences ("He looks for the slippers") • Eye tracking: ∘ Difference score of proportional fixations to object classes in verb window of predictable sentences ∘ Time to direct gaze to named object in predictable and neutral sentences	• Receptive vocabulary • Vocabulary change from experiment to 1.5 years later	• Graded prediction pattern ∘ Increase in anticipatory fixations from low to middle to high predictable objects ∘ Was related with receptive vocabulary (but *not* when controlling for age) • Revision skills: ∘ Longer time to orient gaze to target noun (slippers) in predictable versus neutral sentence. Inaccurate predictions hinder processing ∘ Was related with receptive vocabulary (also when controlling for age) • Word processing speed ∘ Time to fixate on named object in neutral sentences increased with increasing vocabulary (but *not* when controlling for age) • Graded prediction skill and word processing speed were positively related with vocabulary change

(Continues)

TABLE 1 (Continued)

Publication	Age group(s)	Level of processing	Methods and measures	Moderators	Results
Tribushinina and Mak (2016)	• 3 years • Adults	Semantical constraints	• Sentences with attributive versus neutral adjectives and predictable target nouns ("a soft/new pillow") in scene with target (pillow) and distractor (book) object • Eye-tracking: Proportional fixations of target (pillow) in adjective and noun window	-	• 3-year-olds and adults showed more anticipatory target fixation in attributive versus neutral condition. They used semantic cue to predict
Gambi et al. (2016)	• 3–5 years • Young adults	Semantical with structural constraints	• Sentences with constraining verbs ("Jay will *ride* the …") in scenes of two semantic related (horse, cowboy) but only one syntactic-structure related (horse) object • Eye tracking: Proportion of target (horse) fixations in constraining and target window	• Production vocabulary	• Children and adults showed faster and more anticipatory fixations of semantic-and-syntactic related versus semantic related but structural implausible object • No age differences, no influence of productive vocabulary • Children and adults predicted by structure knowledge beyond semantic associations
Sommerfeld et al. (2023)	• 4–5 years • Adults	Semantical constraints	• Sentences with semantically constraining verbs ("The girl eats the cake") in visual scenes of four objects of which either zero, one, three, or four match the verb constraints (are edible), and thus potential prediction options • Eye-tracking: Proportional target fixations in verb and target noun window + target advantage score (to reveal fixation preference for competitors)	• Receptive vocabulary	• Children and adults showed anticipatory fixations of not only one but even three or four visual prediction options. They could predict multiple nouns in parallel by the visuo-linguistic context • Only children's prediction behavior was positively related with receptive vocabulary size
Mani and Huettig (2014)	• 8 years	Semantical constraints	• Same as in Mani & Huettig (2012, see first line)	• Reading skills • Naming tasks • Phonological awareness	• 8-year-olds showed more anticipatory target fixations in constraining versus neutral condition. They used semantic cue to predict • Children with high reading skills showed stronger prediction effect

TABLE 1 (Continued)

Publication	Age group(s)	Level of processing	Methods and measures	Moderators	Results
Smolík and Bláhová (2022)	• 2–2.9 years	Morpho-syntactical constraints	• Sentences with number marked copula ("Where *is/are* the …?") and verbs ("Where *jump/jumps* the …?") in scenes with one versus two objects • Eye tracking: Proportional target fixations in determiner and target window	• Receptive vocabulary • Receptive grammar	• All age groups showed more anticipatory fixations of number matching object upon hearing number marked cue (copula or verb). They used number marking to predict • No effect of receptive vocabulary or grammar
Lukyanenko and Fisher (2016)	• 2.5 years • 3 years • Adults	Morpho-syntactical constraints	• Sentences with number marked copula ("Where *is/are* the …?") in scenes with one versus two objects • Eye-tracking: Proportional target fixations in determiner and target window	-	• 3-year-olds and adults showed more anticipatory fixations of number matching object upon hearing number marked cue. They used number marking to predict
Deevy et al. (2017)	• 3 years	Morpho-syntactical constraints	• Sentences with number marked copula ("*Is/are* the dog running?") in scenes with one versus two objects • Eye tracking: Proportional target fixations in determiner and target window	-	• 3-year-olds showed more anticipatory fixations of number matching object upon hearing number marked cue. They used number marking to predict
Özkan et al. (2022)	• 4–8 years	Morpho-syntactical constraints	• Sentences ("The rabbit eats the carrot") with case marked nouns allowing for the prediction of an upcoming second noun in visual scenes one target and two distractors • Eye tracking: Proportional target fixations in case marked noun and target window	• Large test battery of multiple verbal abilities and capacity of different working memory components	• Children showed more anticipatory fixations of target noun than could be predicted by case marked initial noun • Use of prediction was positively related to productive vocabulary, language production skills, episodic buffer component of working memory
Lindsay et al. (2019)	• 3–5 years • Young adults	Discourse constraints	• Fictional computer game character asks questions with predictive cue early versus late in question • Dependent variable: Answer response time	-	• Children and adults responded faster to questions with early versus late predictive cue • No age differences in using information structure to predict conversation input
Borovsky, Elman, and Fernald (2012)	• 3–10 years • Adults	Discourse constraints	• Participants listen to sentences with agent and action cue and predictable noun ("The *dog chases* the cat while looking at four objects"): agent-and-action-related (cat), agent-related (bone), action-related (treasure), distractor • Eye tracking: Proportional target fixations (cat) in verb and target window	• Receptive vocabulary	• Children and adults anticipatorily fixated on agent-and-action-related object. They combined agent and action cue to predict • Children with high receptive vocabulary showed stronger prediction effect

(Continues)

TABLE 1 (Continued)

Publication	Age group(s)	Level of processing	Methods and measures	Moderators	Results
Borovsky and Creel (2014)	• 3–10 years • Adults	Discourse constraints	• Participants listen to sentence with talker cue (sentence played in voice introduced as pirate) and action cue (hold) and predictable noun ("I want to hold the sword") while looking at four objects: talker-and-action-related (sword), talker-related (ship), action-related (wand), distractor • Eye tracking: Proportional target fixations (sword) in verb window	• Vocabulary size	• Children and adults anticipatorily fixated on talker-and-action-related object. They combined talker and action cue to predict • Larger prediction effect in adults versus children and in participants with higher receptive vocabulary
Lee et al. (2017)	• 7 years • Young adults	Discourse constraints	• Discourse group does, control group does not listen to story that changes sentence constraints ("Fairies love to eat snow"); both groups listen to constraining sentences ("The fairy *eats* the …") while looking at four objects: one semantic (cake) and one story related (snow), two distractors • Eye tracking: Proportional target fixations in verb and target window	• Receptive vocabulary • Inhibitory control • Working memory • Semantic verbal fluency (animals, foods, drinks)	• Children and adults of both groups anticipatorily fixated on semantic related object (snow). Classic prediction effect • Discourse group anticipatorily fixated on story-related object (snow). Children and adults used discourse information to update predictions • Children with higher receptive vocabulary had more problems to override long-term knowledge • Children with lower working memory capacity relied more on prior discourse information (contrary to expectations) • No other significant correlations
Lee et al. (2022)	• 7 years • Young adults	Discourse constraints	• The same as in Lee et al. (2017, see above)	-	• Children and adults of both groups anticipatorily fixated on semantic related object (snow) • Only adults of discourse group anticipatorily fixated on story-related object • Children relied more on real world knowledge during actual online processing

Note: This table provides examples of studies investigating predictive processing with the visual world paradigm in different ages, across age groups, and at different levels of linguistic representations; it does not represent a full review of all studies on this topic.

et al., 2016). For example, prediction error drives learning in recurrent connectionist network models (Elman, 1990). In the dual-path model (Chang et al., 2006), prediction error plays a central role in modeling novel word learning in early development. Similarly, in the predictive interactive multiple memory systems framework, prediction error is associated with the formation of stronger memory traces (Henson & Gagnepain, 2010). Most evidence arguing that prediction error is related to language development comes from correlational studies, which do not allow for conclusions about whether prediction error is the source of learning.

Developmental researchers have started to conduct experimental studies that directly test whether prediction error is a source for learning and memory. Here, we distinguish between two consequences that prediction error may have in the learning phase: (1) immediate updating of linguistic representations, which in turn can facilitate the processing of previously unexpected/disconfirmed linguistic information, and (2) longer-term retrieval from recognition memory when individuals need to retrieve information they have previously encoded in the learning phase (i.e., memory retrieval).

In Table 2, we summarize the main findings of new experimental studies sorted by age and the types of downstream consequence on memory and learning. A few experimental studies have sought to measure the consequences of predictive processing by means of adaptation to changes in frequency of linguistic information, such as syntactic forms (e.g., verb or noun phases; Havron et al., 2019). For instance, in one study (Fazekas et al., 2020), 5- to 6-year-olds who had more exposure to less common dative structures in a learning phase increased their production of dative structures in the subsequent posttest. However, because these studies used frequency manipulations, it is less clear whether children actively used prediction to encode linguistic information in the learning phase. Nevertheless, these types of paradigms are well-suited to systematically vary the amount of exposure during the learning phase, which can help answer this question if one or more repetitions are required in different age ranges to induce learning novel linguistic information.

Other studies have explored how prediction error might support word learning by comparing how children acquired the meanings of (unexpected) novel words in contexts with a stronger versus weaker prediction error (see Table 2). For example, in one study (Reuter et al., 2019), researchers measured the magnitude of prediction error (via eye-tracking) as the difference between preschoolers' initial expectations for a known word and their subsequent shift (redirection) in attention toward a novel item in weakly and highly constraining sentences; novel words appeared in place of highly expected, illustrated nouns (e.g., "Yum! Let's stir the soup with a *cheem*") while the children viewed an image of a (label-known) spoon and a (label-unknown) honey dipper. Preschoolers' subsequent retention of novel words was related to the initial magnitude of this "predict and redirect" response.

But does prediction error affect long-term memory representations? The effects are mixed. While in one study (Reuter et al., 2019), prediction error fostered recognition and learning of novel word meanings in 3- to 5-year-olds, in another (Gambi, Pickering, et al., 2021), researchers found prediction error-driven learning effects in young adults but not in 2- to 4-year-olds. Moreover, in studies using a recognition memory paradigm in young adults, researchers suggested that prediction can also come with downstream memory costs by demonstrating that lexical preactivation of words that were expected but not actually presented can linger in memory and lead to false memory effects in a subsequent word recognition task (e.g., Haeuser & Kray, 2022; Hubbard et al., 2019).

This false prediction-recognition paradigm can directly probe which kind of linguistic information is preactivated and lingers in memory, and support explorations of how this preactivated information changes across development. Such paradigms can be used to determine, for example, whether presented information is maintained in semantic form only or also in phonological form by comparing false alarm rates between semantic and phonological lures of predicted but not presented items in the recognition task. From a developmental perspective, this paradigm would allow researchers to directly test whether children experience more interference than adults from co-activation of phonological lures. In summary, these novel paradigms, which probe the long-term memory consequences of prediction, might illuminate not only when but also under which conditions and what kind of representations children learn.

HOW DO CONSTRAINTS IN LANGUAGE AND COGNITIVE SKILLS AFFECT PREDICTION?

Because age is not an explanatory variable in developmental research, studies on developmental changes in prediction should include measures of language and cognitive skills more frequently. Since predictive processing is likely to be shaped by prior language experience and world knowledge, the magnitude of prediction error should be related to language skills (e.g., Huettig & Janse, 2016). Currently, only correlational evidence supports this view (see Table 1), showing that larger prediction effects emerge in children with larger vocabularies (Borovsky, Elman, & Fernald, 2012; Mani & Huettig, 2012; Özkan et al., 2022; but see Gambi et al., 2016) and more advanced reading skills (Mani & Huettig, 2014).

However, to answer the question of whether language skills such as the size of the lexicon are the source or the consequence of prediction, researchers

TABLE 2 Relations between prediction errors and subsequent learning and memory.

Author/s (year)	Age group(s)	Inducing expectation or prediction (errors) in the learning phase	Effects on learning and recognition memory in the test phase	Results
Ylinen et al. (2017)	• 12 and 24 months	• Familiarization with a syllable sequence (Finnish word *kuko*)	Oddball paradigm: • Presenting unexpected syllables added to familiarized word beginnings, creating either an unexpected novel word (*kuke*) or unexpected unfamiliar word (*kuka*)	• Larger and earlier mismatch responses for unfamiliar and novel word sequences in the familiarized sequence condition • 12- and 24-month-olds used syllable context to recognize familiar and unknown words
Havron, Babineau, and Christophe (2021)	• 18 months	• Familiarization with noun- or verb-related interpretation of novel words	Eye tracking • Gaze monitoring through the experimenter	• No evidence that 18-month-olds in the noun group were expecting a noun or that 18-month-olds in the verb group were expecting a verb
Havron et al. (2019)	• 3–4 years	• Attention to noun or verb phrase by frequency manipulation in the learning phase (noun versus verb groups)	Eye tracking • Proportions of looks to novel actions or unfamiliar objects	• Children in the verb group in the learning phase looked more on a novel action than on novel objects • Children used the distribution of syntactic categories for learning novel meanings • The same effect was found in a young adult control group
Havron, Babineau, Fiévet, et al. (2021)	• 2 years • 3–4 years	• Exposed to predicted nouns versus predicted verbs (less frequent)	Eye tracking • Proportions of looks in verb and noun condition	• 3- to 4-year-olds looked longer at the verb interpretation than at the noun interpretation • This effect was not significant in 2-year-olds
Beretti et al. (2020)	• 4-5-years • Adults	• Exposed to semantic versus syntactic cues in the learning phase	Eye tracking • Predictive looks to object depicting a syntactic or sematic interpretation of novel words	• Predictive looks to the object depicting the familiar structure occurred only in adults, not in 4- to 5-year-olds
Swanson et al. (2022)	• 3–5 years • Adults	• Priming of syntactic structure: *"The girls sleep"* (verb phrase) versus *"The girl's book"* (noun phrase)	Eye tracking • Predictive looks to objects depicting nouns versus actions following novel words	• More predictive looks to the object depicting the primed structure for novel words, in both age groups • Adults and children used their predictions about syntactic structure to infer the meaning of novel words
Fazekas et al. (2020)	• 5–6 years • Adults	• Syntactic • Predictive and unpredictive dative structures	Dative production • Changes in type of dative use from pretest to posttest	• Both age groups showed a shift in the use of less predictable (surprising dative structure) from pretest to posttest • Effect was significant in child group but not in adult group

TABLE 2 (Continued)

Author/s (year)	Age group(s)	Inducing expectation or prediction (errors) in the learning phase	Effects on learning and recognition memory in the test phase	Results
Gambi, Pickering, et al., 2021	• 2–4 years • Adults	• Semantic more/less verb constraining sentences • Combined with familiar/novel objects	• Object choices in the learning phase and the recognition test	• Both age groups were sensitive to the sentence-constraining verbs in their choice of novel objects • Only adults showed better recognition memory for novel objects when they were presented in a highly constraining context (i.e., induced a larger prediction error)
Reuter et al. (2019)	• 3–5 years	• Constrained or unconstrained sentences • Words or pseudowords	Eye tracking • Proportion of target looks in learning and test phase	• Looking behavior in the learning phase to predicted and unpredicted targets was correlated with looking behavior in the test phase but only in constrained sentences • Learning of novel meanings was related to prediction and prediction error (attention redirection)
Borovsky (2022)	• 5–6 years • 7–8 years	• Learning novel event sequences with familiar objects	• Generalization of learned sequence to other event sequences	• Neither 5- to 6-year-olds nor 7- to 8-year-olds learned to predict based on a single sequence exposure but they did so based on multiple ones • 7- to 8-year-olds were more likely to learn from repeated sequences; 5- to 6year-olds learned more from varied sequences
Vergilova et al. (2022)	• 7–13 years	• Weakly or highly constraining sentences • Words or pseudowords	Electroencephalogram • N400 in learning phase (sentence processing) • N400 in the test phase (priming task)	• In the learning phase, N400 to unexpected pseudowords was larger for younger children when the sentence constraint was high • In the test phase, N400 was reduced when primes (words or pseudowords) were shown in a highly constraining sentence; this effect increased with age

Note: This table provides examples of studies investigating relations between prediction error and learning and memory; it does not represent a full review of all studies.

need to use longitudinal designs (see Table 3). Only one study has investigated longitudinal changes in vocabulary size (over 7 months) and its relation to predictive processing (Gambi, Jindal, et al., 2021). In that study, researchers measured predictive processing in a graded fashion by comparing the proportion of fixations on objects that were unpredictable, moderately, and highly predictable, depending on sentence context. The difference between the three conditions (i.e., the graded prediction effect) increased with age from 2 to 5 years, and the increase in vocabulary size after 7 months was correlated with this graded prediction effect. However, since the authors measured only the increase in vocabulary and not changes in the graded prediction effect, the direction of changes between these variables could not be determined. Hence, it remains unclear whether changes in vocabulary in early childhood drive changes in predictive processing or vice versa (see Table 3).

Like linguistic skills, cognitive functioning increases substantially throughout childhood and may support the use of prediction for learning and memory. For example, working memory capacity increases tremendously over childhood (Gathercole et al., 2004) and is tied to other language skills, such as sentence processing (Gilchrist et al., 2009) and reading comprehension (Swanson & Jerman, 2007). Working memory is also linked to other language processing skills. For example, early lexical processing skills at age 2 predict working memory capacity at age 7 (Marchman & Fernald, 2008). In studies of adults, individuals with larger working memory capacity and faster processing speed were more likely to use predictive information (Huettig & Janse, 2016). Furthermore, in a study of 4- to 8-year-olds, working memory capacity was positively correlated with prediction skills (Özkan et al., 2022). These findings suggest that working memory and predictive processing abilities likely influence each other across development.

Recent theoretical assumptions on prediction in adults argue for a tradeoff between memory and surprisal (prediction error). The rationale is that a certain degree of prediction error (by providing more contextual information for prediction) can be achieved only

TABLE 3 Challenges and needs of developmental research on prediction error.

1. *How can intervention (training) studies help infer the causal relations between prediction and language learning?* Intervention studies can simulate longitudinal changes in a shorter time frame by manipulating the degree of exposure to specific aspects of language. Such intervention studies should include a pretest phase to determine baseline performance (e.g., a bias for using a specific lexical structure) and a posttest phase to examine changes from pre- to posttest after a learning phase (exposure to a new lexical structure; cf. Fazekas et al., 2020). Studies of adults have shown that even a single exposure to a new, unknown word in a predictable sentence context can lead to fast learning of novel word meanings (Borovsky et al., 2010; Borovsky, Elman, & Kutas, 2012). Studies with young children have demonstrated that they can infer word meanings from context in "fast-mapping" paradigms with limited alternative options and strong linguistic and referential support. However, it seems that children's ability to engage in one-shot learning from sentence context alone increases throughout development (Vergilova et al., 2022). Researchers need to address under which conditions one-shot learning is supported in early childhood. In general, we do not known how much exposure (practice) is needed at different levels of linguistic encodings throughout early development for integration of information in short- and long-term memory. Compared to adults, children likely need more exposure to novel unknown words before those words can be integrated into their semantic knowledge base (Borovsky, 2022).
2. *What is the direction of developmental changes between prediction (error) and learning and memory*? The field needs more longitudinal studies to determine more effectively the direction of these relations throughout childhood development (cf. Rabagliati et al., 2016). Although preliminary evidence suggests relations between prediction and language learning, the direction of developmental changes has not been investigated. Longitudinal research designs would allow studies to determine whether prediction (error) can explain developmental changes in language learning, such as fostering an increase in vocabulary size. Alternatively, developmental changes in the knowledge base (e.g., vocabulary size) could be the determiner for developmental changes in using prediction during language processing. We speculate that both skills are likely to play a role, but that the relative importance of these relations will vary with age. In terms of methods to test the importance of the direction of developmental changes (e.g., prediction error → lexicon; lexicon → prediction error), latent change growth models are particularly suitable. These models also allow researchers to test whether the change depends on initial performance level (e.g., lexicon size) and whether the change is correlated to changes in other domains of functioning, such as cognition (e.g., working memory). If prediction is measured at two different levels of representation (syntactic and semantic), such models can also test how changes at these levels are related to each other.
3. *How do capacity limitations in language and cognitive skills during early childhood contribute to using prediction (error) and its relations to learning and memory?* Among the critical questions are: To what extent do linguistic skills or domain-general cognitive functions support prediction? Do children leverage these skills to predict upcoming input, or does prediction enable the maturation of these cognitive functions? Age differences in using prediction (error) may be mediated by individual differences in language and cognitive skills. For instance, prediction (error) may occur only with a certain level of world knowledge. If the knowledge base is small, the likelihood of inducing a prediction error is also low. Also, working memory capacity is needed to maintain relevant semantic information until the end of a sentence to produce a prediction error that could limit capacity in early childhood. Although some evidence suggests that individual differences in vocabulary size, reading skills, and processing speed are related to prediction errors, the general pattern of results is mixed. One reason may be the lack of reliability of measurement because researchers used only single indicator tests. Therefore, researchers should create a battery of tests to measure language skills (e.g., productive and receptive vocabulary, verbal and reading fluency) and cognitive skills (e.g., working memory, speed of processing, inhibition), ideally at the construct level (with more than one indicator test) to increase the reliability and validity of measurement. Doing so will help researchers identify which capacity limitations in early childhood contribute to using prediction error and its relations to learning and memory. Conducting data of this sort and measuring multiple tests will be challenging in early infancy.

by storing previous linguistic information in memory (Hahn et al., 2021). Some (e.g., Futrell et al., 2020) also claim that processing difficulties elicited by a word in context depend on "lossy" memory representations of the previous context (i.e., a potentially incomplete contextual representation). Because working memory capacity increases substantially throughout childhood, these cognitive constraints may modulate engagement in linguistic prediction depending on the complexity of the task (e.g., working memory may be more engaged at the discourse level than at the lexical level). More evidence is needed to help substantiate this view. Although the assumption of memory is implemented in neural networks about language learning in childhood, cognitive constraints such as incomplete memory representations, among others, are not, even though they may help explain why prediction (error) and its impact on learning and memory change with age.

In summary, researchers should include individual differences in language and cognitive skills as constraining variables in studies of developmental changes in prediction. They should also conduct more longitudinal studies to determine the direction of changes (e.g., prediction → lexicon size, or lexicon size → prediction). (For more information about the challenges and needs of forthcoming research, see Table 3.)

CONCLUSION

Children use prediction to anticipate upcoming linguistic information, but evidence on longitudinal developmental changes at different levels of linguistic representations and our understanding of how they inform each other is still new. In this emerging research field, developmental studies can provide data about the role of prediction error in updating, learning, and memorizing linguistic information in childhood. Paradigms that manipulate the frequency of linguistic information are well-suited to help researchers examine the amount of exposure needed for learning, while paradigms probing the long-term memory consequences of prediction can determine under which conditions and what kind of representations children learn. Researchers need to undertake longitudinal studies and use multiple measures of language and cognitive skills to answer the questions of at which age, under which conditions, and why prediction aids language learning.

ACKNOWLEDGMENT
Open Access funding enabled and organized by Projekt DEAL.

FUNDING INFORMATION
Funding for the preparation of this article came from the German Research Foundation, "Information Density and Linguistic Encoding" Project-ID 232722074–SFB1102.

ORCID
Jutta Kray https://orcid.org/0000-0002-2119-9042

REFERENCES

Babineau, M., Havron, N., Dautriche, I., de Carvalho, A., & Christophe, A. (2023). Learning to predict and predicting to learn: Before and beyond the syntactic bootstrapper. *Language Acquisition*, *30*(3–4), 337–360. https://doi.org/10.1080/10489223.2022.2078211

Beretti, M., Havron, N., & Christophe, A. (2020). Four-and 5-year-old children adapt to the reliability of conflicting sources of information to learn novel words. *Journal of Experimental Child Psychology*, *200*(December), 104927. https://doi.org/10.1016/j.jecp.2020.104927

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

Borovsky, A. (2022). Developmental changes in how children generalize from their experience to support predictive linguistic processing. *Journal of Experimental Child Psychology*, *219*(July), 105349. https://doi.org/10.1016/j.jecp.2021.105349

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

Borovsky, A., Elman, J. L., & Fernald, A. (2012). Knowing a lot for one's age: Vocabulary skill and not age is associated with anticipatory incremental sentence interpretation in children and adults. *Journal of Experimental Child Psychology*, *112*(4), 417–436. https://doi.org/10.1016/j.jecp.2012.01.005

Borovsky, A., Elman, J. L., & Kutas, M. (2012). Once is enough: N400 indexes semantic integration of novel word meanings from a single exposure in context. *Language Learning and Development*, *8*(3), 278–302. https://doi.org/10.1080/15475441.2011.614893

Borovsky, A., Kutas, M., & Elman, J. (2010). Learning to use words: Event-related potentials index single-shot contextual word learning. *Cognition*, *116*(2), 289–296. https://doi.org/10.1016/j.cognition.2010.05.004

Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review*, *113*(2), 234–272. https://doi.org/10.1037/0033-295X.113.2.234

Cholewa, J., Neitzel, I., Bürsgens, A., & Günther, T. (2019). Online-processing of grammatical gender in noun-phrase decoding: An eye-tracking study with monolingual German 3rd and 4th graders. *Frontiers in Psychology*, *10*, 2586. https://doi.org/10.3389/fpsyg.2019.02586

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181–204. https://doi.org/10.1017/S0140525X12000477

de Carvalho, A., Dautriche, I., Fiévet, A. C., & Christophe, A. (2021). Toddlers exploit referential and syntactic cues to flexibly adapt their interpretation of novel verb meanings. *Journal of Experimental Child Psychology*, *203*(March), 105017. https://doi.org/10.1016/j.jecp.2020.105017

Deevy, P., Leonard, L. B., & Marchman, V. A. (2017). Sensitivity to morphosyntactic information in 3-year-old children with typical language development: A feasibility study. *Journal of Speech, Language, and Hearing Research*, *60*(3), 668–674. https://doi.org/10.1044/2016_JSLHR-L-15-0153

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, *14*(2), 179–211. https://doi.org/10.1207/s15516709cog1402_1

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. https://doi.org/10.1098/rsos.180877

Friederici, A. D. (2005). Neurophysiological markers of early language acquisition: From syllables to sentences. *Trends in Cognitive Sciences*, *9*(10), 481–488. https://doi.org/10.1016/j.tics.2005.08.008

Friedrich, M., & Friederici, A. D. (2005). Semantic sentence processing reflected in the event-related potentials of one-and two-year-old children. *Neuroreport*, *16*(16), 1801–1804. https://doi.org/10.1097/01.wnr.0000185013.98821.62

Friston, K. J. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127–138. https://doi.org/10.1038/nrn2787

Futrell, R., Gibson, E., & Levy, R. P. (2020). Lossy-context surprisal: An information-theoretic model of memory effects in sentence processing. *Cognitive Science*, *44*(3), e12814. https://doi.org/10.1111/cogs.12814

Gambi, C., Jindal, P., Sharpe, S., Pickering, M. J., & Rabagliati, H. (2021). The relation between preschoolers' vocabulary development and their ability to predict and recognize words. *Child Development*, *92*(3), 1048–1066. https://doi.org/10.1111/cdev.13465

Gambi, C., Pickering, M. J., & Rabagliati, H. (2016). Beyond associations: Sensitivity to structure in pre-schoolers' linguistic predictions. *Cognition*, *157*(December), 340–351. https://doi.org/10.1016/j.cognition.2016.10.003

Gambi, C., Pickering, M. J., & Rabagliati, H. (2021). Prediction error boosts retention of novel words in adults but not in children. *Cognition*, *211*(June), 104650. https://doi.org/10.1016/j.cognition.2021.104650

Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental Psychology*, *40*(2), 177–190. https://doi.org/10.1037/0012-1649.40.2.177

Gilchrist, A. L., Cowan, N., & Naveh-Benjamin, M. (2009). Investigating the childhood development of working memory using sentences: New evidence for the growth of chunk capacity. *Journal of Experimental Child Psychology*, *104*(2), 252–265. https://doi.org/10.1016/j.jecp.2009.05.006

Günther, T., Kirschenkern, A., Mayer, A., Steinke, F., & Cholewa, J. (2023). The development of predictive gender processing strategies during noun phrase decoding: An eye-tracking study with German 5-to 10-year-old children. *Journal of Speech, Language, and Hearing Research*, *66*(10), 1–18. https://doi.org/10.1044/2023_JSLHR-23-00100

Haeuser, K. I., & Kray, J. (2022). How odd: Diverging effects of predictability and plausibility violations on sentence reading and word memory. *Applied PsychoLinguistics*, *43*(5), 1193–1220. https://doi.org/10.1017/S0142716422000364

Hahn, M., Degen, J., & Futrell, R. (2021). Modeling word and morpheme order in natural language as an efficient trade-off of memory and surprisal. *Psychological Review*, *128*(4), 726–756. https://doi.org/10.31234/osf.io/nu4qz

Havron, N., Babineau, M., & Christophe, A. (2021). 18-month-olds fail to use recent experience to infer the syntactic category of novel words. *Developmental Science*, *24*(2), e13030. https://doi.org/10.1111/desc.13030

Havron, N., Babineau, M., Fiévet, A. C., de Carvalho, A., & Christophe, A. (2021). Syntactic prediction adaptation accounts for language processing and language learning. *Language Learning*, *71*(4), 1194–1221. https://doi.org/10.1111/lang.12466

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

Henson, R. N., & Gagnepain, P. (2010). Predictive, interactive multiple memory systems. *Hippocampus*, *20*(11), 1315–1326. https://doi.org/10.1002/hipo.20857

Hubbard, R. J., Rommers, J., Jacobs, C. L., & Federmeier, K. D. (2019). Downstream behavioral and electrophysiological consequences of word prediction on recognition memory. *Frontiers in Human Neuroscience*, *13*, 291. https://doi.org/10.3389/fnhum.2019.00291

Huettig, F., & Janse, E. (2016). Individual differences in working memory and processing speed predict anticipatory spoken language processing in the visual world. *Language, Cognition and Neuroscience*, *31*(1), 80–93. https://doi.org/10.1080/23273798.2015.1047459

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

Kuperberg, G. R., & Jaeger, T. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, *31*(1), 32–59. https://doi.org/10.1080/23273798.2015.1102299

Kutas, M., Federmeier, K. D., & Urbach, T. P. (2014). The "negatives" and "positives" of prediction in language. In M. S. Gazzaniga & G. R. Mangun (Eds.), *The cognitive neurosciences* (pp. 649–656). Boston Review.

Lee, R., Chambers, C. G., Huettig, F., & Ganea, P. A. (2017). Children's semantic and world knowledge overrides fictional information during anticipatory linguistic processing. In G. Gunzelmann, A. Howes, T. Tenbrink, & E. Davelaar (Eds.), *The 39th annual meeting of the Cognitive Science Society* (pp. 730–735). Cognitive Science Society.

Lee, R., Chambers, C. G., Huettig, F., & Ganea, P. A. (2022). Children's and adults' use of fictional discourse and semantic knowledge for prediction in language processing. *PLoS One*, *17*(4), e0267297. https://doi.org/10.1371/journal.pone.0267297

Lindsay, L., Gambi, C., & Rabagliati, H. (2019). Preschoolers optimize the timing of their conversational turns through flexible coordination of language comprehension and production. *Psychological Science*, *30*(4), 504–515. https://doi.org/10.1177/0956797618822802

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

Mani, N., & Huettig, F. (2012). Prediction during language processing is a piece of cake—But only for skilled producers. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(4), 843–847. https://doi.org/10.1037/a0029284

Mani, N., & Huettig, F. (2014). Word reading skill predicts anticipation of upcoming spoken language input: A study of children developing proficiency in reading. *Journal of Experimental Child Psychology*, *126*(October), 264–279. https://doi.org/10.1016/j.jecp.2014.05.004

Marchman, V. A., & Fernald, A. (2008). Speed of word recognition and vocabulary knowledge in infancy predict cognitive and language outcomes in later childhood. *Developmental Science*, *11*(3), F9–F16. https://doi.org/10.1111/j.1467-7687.2008.00671.x

Özkan, D., Küntay, A. C., & Brouwer, S. (2022). The role of verbal and working memory skills in Turkish-speaking children's morphosyntactic prediction. *Applied PsychoLinguistics*, *43*(6), 1305–1328. https://doi.org/10.1017/S0142716422000388

Parise, E., & Csibra, G. (2012). Electrophysiological evidence for the understanding of maternal speech by 9-month-old infants. *Psychological Science*, *23*(7), 728–733. https://doi.org/10.1177/0956797612438734

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

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

Reuter, T., Borovsky, A., & Lew-Williams, C. (2019). Predict and redirect: Prediction errors support children's word learning.

Developmental Psychology, *55*(8), 1656–1665. https://doi.org/10.1037/dev0000754

Smolík, F., & Bláhová, V. (2022). Here come the nouns: Czech two-year-olds use verb number endings to predict sentence subjects. *Cognition*, *219*(February), 104964. https://doi.org/10.1016/j.cognition.2021.104964

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*(November), 105690.

Swanson, E., Frank, M. C., & Degen, J. (2022). Syntactic adaptation and word learning in children and adults. *PsyArXiv*. https://doi.org/10.31234/osf.io/zkvcq

Swanson, H. L., & Jerman, O. (2007). The influence of working memory on reading growth in subgroups of children with reading disabilities. *Journal of Experimental Child Psychology*, *96*(4), 249–283. https://doi.org/10.1016/j.jecp.2006.12.004

Tribushinina, E., & Mak, W. M. (2016). Three-year-olds can predict a noun based on an attributive adjective: Evidence from eye-tracking. *Journal of Child Language*, *43*(2), 425–441. https://doi.org/10.1017/S0305000915000173

Vergilova, Y., Jachmann, T. K., Mani, N., & Kray, J. (2022). Age-related differences in expectation-based novel word learning. *Psychophysiology*, *59*(8), e14030. https://doi.org/10.1111/psyp.14030

Yazbec, A., Kaschak, M. P., & Borovsky, A. (2019). Developmental timescale of rapid adaptation to conflicting cues in real-time sentence processing. *Cognitive Science*, *43*(1), e12704. https://doi.org/10.1111/cogs.12704

Ylinen, S., Bosseler, A., Junttila, K., & Huotilainen, M. (2017). Predictive coding accelerates word recognition and learning in the early stages of language development. *Developmental Science*, *20*(6), e12472. https://doi.org/10.1111/desc.12472

Zettersten, M. (2019). Learning by predicting: How predictive processing informs language development. In B. Busse & R. Moehlig-Falke (Eds.), *Patterns in language and linguistics* (pp. 255–288). de Gruyter.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Kray, J., Sommerfeld, L., Borovsky, A., & Häuser, K. (2024). The role of prediction error in the development of language learning and memory. *Child Development Perspectives*, *18*, 190–203. https://doi.org/10.1111/cdep.12515