<https://helda.helsinki.fi>

Predictive coding accelerates word recognition and learning in the early stages of language development

Ylinen, Sari

2017-11

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. , vol. 20 , no. 6 , 12472 . <https://doi.org/10.1111/desc.12472>

<http://hdl.handle.net/10138/229888>

<https://doi.org/10.1111/desc.12472>

acceptedVersion

Downloaded from Helda, University of Helsinki institutional repository.

This is an electronic reprint of the original article.

This reprint may differ from the original in pagination and typographic detail.

Please cite the original version.

Title: Predictive coding accelerates word recognition and learning in the early stages of language development

Running head: Predictions in children's word recognition and learning

Authors: S. Ylinen^{a*}, A. Bosseler^a, K. Junttila^a, and M. Huotilainen^{a,b}

Affiliations: ^aCognitive Brain Research Unit, Institute of Behavioural Sciences, University of Helsinki, PO Box 9, FIN-00014 University of Helsinki, Finland.

^bFinnish Institute of Occupational Health, Topeliuksenkatu 41 a A, FIN-00250 Helsinki, Finland.

***Correspondence to:** Sari Ylinen, Cognitive Brain Research Unit, Institute of Behavioural Sciences, P.O. Box 9, FIN-00014 University of Helsinki, Finland; Phone: +358 50 4484 170; e-mail: sari.ylinen@helsinki.fi

Acknowledgments: The authors thank the participants and their families, Miika Leminen and Tommi Makkonen for technical assistance, Jenni Väre and Leena Wallendahr for assistance with the recordings, and two anonymous reviewers for constructive comments on the manuscript. This work was supported by the Academy of Finland (grant numbers 131963, 274058, and 268231).

Highlights:

- The first two years of a child's life is a period of rapid vocabulary growth, yet the neural processes involved in word learning are not fully known.
- Our data recorded with electroencephalography suggest that word recognition in the child brain is accelerated by predictive inference, which represents a domain-general principle of information processing in the brain.
- At 12 months, the strength of infants' brain responses to unexpected novel words are correlated with their vocabulary scores. This suggests that predictive inference, particularly the brain response signaling prediction error, is linked with word learning.
- Predictive coding may mediate attention allocation in young children, by means of which it may drive the learning of both word forms and the link to their referents.

Abstract: The ability to predict future events in the environment and learn from them is a fundamental component of adaptive behavior across species. Here we propose that inferring predictions facilitates speech processing and word learning in the early stages of language development. Twelve- and 24-month olds' electrophysiological brain responses to heard syllables are faster and more robust when the preceding word context predicts the ending of a familiar word. For unfamiliar, novel word forms, however, word-expectancy violation generates a prediction error response, the strength of which significantly correlates with children's vocabulary scores at 12 months. These results suggest that predictive coding may accelerate word recognition and support early learning of novel words, including not only the learning of heard word forms but also their mapping to meanings. Prediction error may mediate learning via attention, since infants' attention allocation to entire learning situation in natural environments could account for the link between prediction error and the understanding of word meanings. On the whole, the present results on predictive coding support the view that principles of brain function reported across domains in humans and non-human animals apply to language and its development in the infant brain.

Introduction

In both humans and non-human animals, the capacity to predict future events on the basis of probabilistic inference is crucial for adaptation to the environment and, in some cases, even for survival. Predictive coding theory (Rao & Ballard, 1999; Friston, 2005, 2009, 2010) sheds light on the neural underpinnings of such predictions. According to the theory, predictive coding is an implicit process that creates an internal model about sensory input with the aim of minimizing surprise. It is based on bidirectional information flow in a hierarchical neural network. The inferred causes of the incoming sensory input are encoded by representational units located at a higher level of the network. They send predictions to prediction error units located at a lower level, where bottom-up input is compared with these top-down predictions. A match between input and prediction results in a suppressed neural response, whereas a mismatch elicits a prediction error response, which is projected back to the higher level to adjust the internal model. As a result, only input that does not match the internal model requires further processing resources.

As recently reviewed by Pouget, Beck, Ma, and Latham (2013), the brain uses probabilistic inference to solve many tasks, ranging from sensory processing that is shared by species to higher cognitive functions that are specific to humans. One of the neurocognitive processes observed across species and linked with probabilistic inference is the elicitation of the mismatch negativity (MMN) component of auditory event-related potential (ERP) (Näätänen, Gaillard, & Mäntysalo, 1978; for reviews on MMN, see Näätänen, Paavilainen, Rinne, & Alho, 2007; Winkler, 2007; Näätänen, Astikainen, Ruusuvirta, & Huotilainen, 2010). MMN is typically elicited in an oddball paradigm, which refers to a stimulus sequence with a high probability of repetitive *standard* stimuli and occasional *deviant* stimuli. MMN reflects automatic auditory processing in the brain, since it is elicited even when participant's attention is directed away from the sounds. Winkler (2007) has

proposed that the brain uses the regularities of the auditory environment to create a model that predicts the sounds that will follow. Violations of these predictions elicit the MMN, which is in line with the predictive coding theory. Furthermore, Friston (2005) has explicitly proposed that the MMN belongs to the family of prediction error responses (see also Wacongne et al., 2011).

Interestingly, recent studies have suggested that the domain-general principles of predictive coding seem to also serve the processing of language. Gagnepain, Henson, and Davis (2012) have proposed that human adults use predictive coding to recognize spoken words. Specifically, their pattern of findings is compatible with the interpretation that compared with brain responses to unpredicted speech sounds, speech sounds predicted on the basis of long-term memory word representations generate reduced responses in superior temporal cortex. In contrast, speech sounds that do not match word-based predictions generate enhanced brain responses, interpreted as prediction error signals. Thus, predictive coding may, together with several other factors, contribute to speech recognition. However, the principle of predictive coding has not been associated with recognition only, but also with learning, because predictions promote reward-based associative learning in non-human animals (Schultz, Dayan, & Montague, 1997; Steinberg et al., 2013). Bridging the gap between these findings, we postulate that predictive coding could support both spoken-word recognition and word learning in infants and young children, as they show the ability of predictive coding from birth (Trainor, 2012; Haden, Nemeth, Torok, & Winkler, 2015).

To test this hypothesis with recordings of auditory event-related potential (ERP), we presented 12- and 24-month-old children with native language syllables in an oddball paradigm, where syllable patterns were expected to generate predictions at different

levels. These predictions can be probed by violating them, which elicits the mismatch response (child equivalent of the MMN)¹ or prediction error. The hypothesis on different levels of prediction is based on previous studies that have, on the one hand, suggested that MMN reflects hierarchical predictions derived from probabilities within sound sequences (Wacongne et al., 2011; Basirat, Dehaene, & Dehaene-Lambertz, 2014) and, on the other hand, shown that predictions and consequent prediction error responses may also be driven by word representations in the long-term memory (Gagnepain et al., 2012). Thus, when an oddball paradigm includes isolated syllables, they most likely induce predictions about the following syllables on the basis of stimulus probabilities in the stimulus sequence. Violations of these sequence-level predictions about the following items are expected to elicit mismatch responses. In contrast, when the same syllables are presented within words as second syllables, the word beginnings may induce predictions about these syllables at the level of word representations. In this case, the violations of predictions are expected to generate prediction error responses at word level.

In addition to probing predictions with their violations, the oddball paradigm can be used to study the activation of long-term memory representations for words (for a review, see Pulvermüller & Shtyrov, 2006). When presented among pseudowords, words typically elicit larger MMN responses than pseudowords, because deviant syllables completing words activate the corresponding word representations resulting in enhanced responses (Pulvermüller et al. 2001; Garagnani, Wennekers, & Pulvermüller, 2008; Garagnani & Pulvermüller, 2011). According to the neurocomputational model by Garagnani and Pulvermüller (2011), local inhibition of cortical activity or the combination of inhibition and neuronal adaptation may account for MMN responses in non-speech tones, but not in

¹ Since in infants and young children the brain responses elicited in an oddball paradigm and resembling MMN with respect to their function can be negative or positive in polarity, we will further use the terms mismatch response to refer to this response in children and MMN to refer to it in adults.

words. The most likely explanation for their pattern of results in words is the “ignition” and reverberation of activity in the neural network covering distant brain areas, including areas involved in speech perception and production (Garagnani & Pulvermüller, 2011, see also Pulvermüller, 1999).

To reveal the effect of word-level predictions and lexical activations in a setup controlling for the effect of sequence-level predictions and neuronal adaptation in repetitive stimulation, we compared ERP responses in the two conditions that used the same syllables of interest. Firstly, in No context condition (Fig. 1A), we presented the standard syllable, occurring with a high probability, in isolation. Secondly, in Context condition (Fig. 1B), we presented the same standard syllable after a familiar word beginning, enabling to interpret the syllable of interest as the word ending. Thus, the critical difference between the two conditions was the absence or presence of the context syllable. In both conditions, the standard syllable was expected to create sequence-level predictions of hearing this particular syllable (Fig. 1C and D, longer arrows). In Context condition, however, the context syllable representing a familiar word beginning was expected to create predictions at another level, namely, to create word-level predictions of hearing the ending of some familiar word (Fig. 1D, shorter arrows; for discussion on knowledge-based predictions, see Poeppel & Monahan, 2011; for phonological predictions in pseudowords, see Ylinen et al., submitted). To probe the predictions at sequence and word levels, we presented two deviant syllables (Fig. 1A and B). In No context condition, both deviants were equally unexpected with respect to sequence-level predictions and hypothesized to elicit typical mismatch responses (see Table 1). In Context condition, however, the same deviant syllables either completed a familiar word, the meaning of which the children knew, or a novel word, the meaning of which the children could not know (Fig. 1B). The involvement

of words in Context condition was hypothesized to enable the interaction of sequence-level and word-level predictions and to result in different brain responses to the deviant syllables (see Table 1): The deviant syllable completing the familiar word that was not the most probable in the sequence was expected to activate the lexical representation for this word (Pulvermüller et al. 2001; Garagnani et al., 2008; Garagnani & Pulvermüller, 2011). In contrast, the deviant syllable completing the novel word was expected to result in a prediction error with respect to lexical candidates (Gagnepain et al., 2012). Efficient processing of novelty as reflected by the prediction error might be linked with children's ability to learn novel words, resembling the link between the prediction error and learning in animals (Schultz et al., 1997; Steinberg et al., 2013).

We hypothesized that 1) if predictive coding supports word recognition in children, then the brain responses to syllables completing familiar words will be faster and more robust than those to isolated syllables that are not predicted by the context. This is because hearing word beginnings may enable top-down lexical predictions to participate in the encoding of auditory input (Gagnepain et al., 2012), resulting in accelerated and facilitated processing of the final syllable. We also hypothesized that 2) if predictive coding supports the learning of new words, then the strength of the prediction error responses for syllables completing novel words will correlate with measures of vocabulary development. In addition, 3) the responses to syllables completing novel words are hypothesized to be larger and faster than the responses to the same syllables in isolation that are not predicted by the context. This is because the novel word violating top-down word-level predictions was expected to elicit a prediction error response with respect to predictions both at word and sequence levels.

Materials and Methods

Participants

The ages between 12 to 24 months span the period from which children are learning to produce their first words to the period when they show a marked increase in the number of words produced. Therefore, the present study focused on these ages to observe possible differences along language development or maturation. Twenty-one 12-month-old (± 1 mo) and twenty 24-month-old (± 1 mo) children living in Finnish-speaking families² were recruited for the electroencephalography (EEG) measurements. Data of some children were, however, excluded due to familial risk for language-related problems, delayed language development diagnosed after the recording, technical problems, or not reaching the criterion of 19 accepted epochs (for a similar criterion, see Conboy & Kuhl, 2011; Basirat et al., 2014). As a result, ten 12-month-olds (4 girls and 6 boys) and fourteen 24-month-olds (4 girls and 10 boys) were included in the analysis. According to parental reports, these children were born full-term, were healthy, developed normally, and were likely to have no speech or language-related dysfunctions. Although the number of participants in each group is relatively small, it is noteworthy that the main findings of the study are based on the data of all 24 children.

Stimuli and study design

Stimulus material consisted of syllables [ko], [kɑ], and [ke], which were the stimuli of interest, and a context syllable [ku]. The stimuli were constructed as follows. During recording in an acoustically shielded chamber, a female native Finnish speaker pronounced the Finnish words /kuk:o/ and /kuk:a/ and the phonotactically legal pseudoword /kuk:e/ several times. Thirty best exemplars of each syllable [ku], [ko], [kɑ], and [ke] were chosen (altogether 120 syllables). For the context syllables [ku], the mean

² The children were estimated not to have disadvantaged socio-economical status.

duration was 145 (± 9) ms, the mean intensity was 81 (± 1) dB, and the mean pitch 253 (± 5) Hz. The mean formant frequencies F1-F3 were 472, 761, and 2713 Hz, respectively. For the syllables [ko], [kɑ], and [ke], the mean durations were 170 (± 17), 178 (± 23), and 172 (± 15) ms, mean pitch was 211 (± 7), 208 (± 6), and 209 Hz (± 6), respectively, and the mean intensity was 79 (± 2) dB. The acoustic variability was thus larger within each stimulus type than between them. On average, the formant frequencies F1-F3 were 584, 1226, and 2762 Hz for [ko], 642, 1422, and 2657 Hz for [kɑ], and 515, 1825, and 2682 for [ke], respectively.

The study design included two conditions. Both conditions used an oddball paradigm with frequent standard and infrequent deviant stimuli, presented pseudorandomly (there were at least two standards between deviants). In No context condition (Fig. 1 A), where syllables of interest were presented in isolation, syllable [ko] served as the standard stimulus and was presented with a probability of 80 %. Syllables [ke] and [kɑ] served as the deviant stimuli, each of which was presented with a 10 % probability. Context condition, in turn, included the same syllables, but they were always preceded by the context syllable [ku] (Fig. 1 B). In this condition, the standard stimulus was [kuk:o] 'rooster', consisting of the context syllable [ku] and the following syllable [ko] and presented with a probability of 80 %. In addition, two deviants were introduced, each with a 10 % probability. The first deviant was [kuk:e], consisting of the context syllable [ku] and the following syllable [ke]. [kuk:e] could not be familiar to the children, because it was a pseudoword (i.e., it was a phonologically legal word form with no meaning in the Finnish language). For simplicity, [kuk:e] is hereafter referred to as a novel word (in practice, it did not differ from other novel words the children encounter and the meaning of which they do not know). The second deviant was the word [kuk:a] 'flower', consisting of the context syllable [ku] and the following syllable [kɑ]. This word was familiar to the children.

To construct stimuli for Context condition, the context syllables [ku] were chosen from different words/pseudowords and randomly paired with the syllables of interest. This was to demonstrate that the recognition of the syllables of interest could not be determined by any acoustical cues of the context syllables. Specifically, because of the intentional random pairing of the syllables, possible co-articulatory cues of the context syllables could not have consistently predicted the final syllables, and the effect of co-articulation, if any, was the same for all stimulus types³. Thus, possible predictions derived from the context syllables must be based on long-term memory representations for words rather than acoustical features. The context syllable [ku] was always followed by a 200 ms silent period, mimicking the occlusion phase of the geminate stop consonant /k:/, and then one of the syllables of interest (Fig. 1 B).

Context and No context conditions were presented in separate stimulus blocks. Each block consisted of 250 stimuli. Context and No context blocks were run in alternating fashion until the child became restless or the caregiver wished to terminate the experiment (maximum six blocks per condition). In both conditions, the inter-stimulus interval was 600 ms.

Procedure and data analysis

Two weeks before the experiment, participants received a picture sheet that included eight pictures of various objects, mostly common animals. The objects were named in the sheet to ensure that the objects were named in the same way to all children. The pictures and

³ Note that the major source of co-articulation, the consonant /k:/, was the same for all stimuli. Acoustical measurements conducted in a previous study (Ylinen et al., submitted) suggested that in similar bisyllabic Finnish stimuli, first-syllable vowels had a minimal co-articulatory effect on following vowels, not exceeding just noticeable difference.

names included those of a rooster (in Finnish *kukko*) and a flower (in Finnish *kukka*), representing the words of interest in the present study (note that *kukke*, the novel word, was not included). Parents were instructed to look at the picture sheet together with a child several times so that each picture would be named altogether 6-10 times. They were instructed to point to each picture and simultaneously name it aloud. They were also encouraged to ask questions such as 'where is a flower?' from children to gather information about their understanding of the words. Within one week prior to the measurement, the parents also filled in the Finnish version (Lyytinen, 1999) of MacArthur-Bates communicative development inventory (CDI; Fenson et al., 1994) and a questionnaire about medical condition and general development of their child. CDI for 12-month-olds contained separate assessment of receptive and productive vocabulary, but since most of our 12-month-olds produced 0-2 words, receptive vocabulary was chosen to index vocabulary development. CDI for older children contained only the assessment of productive vocabulary.

During the EEG measurement, children sat on their parent's lap in an acoustically and electrically shielded room and watched a silent cartoon of choice from a screen placed in front of them. Stimuli were presented via loudspeakers located on both sides of the screen with a comfortable listening level (about 54 dB at the midpoint of the two loudspeakers). Electrodes were attached to F3, F4, C3, C4, Pz, the two mastoids, and below the right eye. EEG was measured with Neuroscan system and Synamps amplifier using 500 Hz sampling rate, 0.1-70 Hz filter, and Fpz as a reference. After the recording, the data were re-referenced to the average of the two mastoids and filtered using a pass-band of 1-30 Hz (slope 24 dB/octave). Artifacts exceeding 120 μ V were rejected. Epochs of -100-600 ms with respect to the beginnings of the syllables of interest were averaged for each stimulus

type (two standards following deviants were omitted from averaging). Baseline correction (-100-0 ms with respect to the beginnings of the syllables of interest) was applied to all waveforms. To quantify mismatch responses with a minimal contribution of exogenous ERP components linked to sound properties, responses to standards were subtracted from those to deviants. Mean amplitudes were measured from individuals' deviant-minus-standard difference waveforms at fronto-central sites in four time windows: 40-100, 130-190, 190-250, and 310-370 ms, determined on the basis of previous literature and the observation of differences between stimulus types in grand-average waveforms. Although being wider, these time windows, following the lexical divergence point (~20 ms from the second syllable onset) by 20-80 ms and 110-170 ms, were chosen to resemble those used in adults by MacGregor, Pulvermüller, van Casteren, & Shtyrov (2012) and Pulvermüller et al. (2001), who observed lexical effects in electromagnetic responses at 50-80 ms and 110-170 ms from the word recognition point (MacGregor et al., 2012) or 150-180 ms from critical syllable onset (Pulvermüller et al., 2001). By choosing time windows at similar latencies, our intention was to enable rough comparison between children's and adults' data. In addition to studying difference waveforms, ERP mean amplitudes were measured from standard responses in the earliest time window of 40-100 ms to correlate vocabulary scores with possible ERP suppression effects that could, in part, account for the effects observed in difference waveforms.

Statistical tests

The mean ERP amplitudes at fronto-central sites were computed from individuals' difference waveforms and submitted to 2 x 2 x 2 x 2 x 2 repeated measures analysis of variance (ANOVA) with between-subject variable Group (12 mo vs. 24 mo) and within-subject variables Context (word context vs. no context), Familiarity ([kɑ] ending a familiar

word vs. [ke] ending a novel word in Context condition, but both syllables being equally familiar in No context condition), Coronal scalp site (frontal vs. central) and Hemisphere (left vs. right). ANOVAs were run separately for each time window. Significant interactions were followed up using Bonferroni-corrected pairwise comparisons and Cohen's *d* effect sizes were calculated for each paired comparison. Since we hypothesized that the strength of prediction error responses might be linked with vocabulary development, we tested correlations between the mean amplitudes of prediction error responses and CDI scores. They were determined using Robust correlation toolbox for Matlab (Pernet, Wilcox, & Rousselet, 2013) to minimize the effect of outliers. Robust_correlation.m function of the toolbox was used to ensure that the data were normally distributed and to identify outliers in the data. Since the data were normally distributed but bivariate outliers were found, we used Pearson's *r* skipped correlation, which down-weights or removes outliers and accounts for them in significance testing (Rousseeuw, 1984; Rousseeuw & Van Driessen, 1999; Verboten & Hubert, 2005; Pernet, Wilcox, & Rousselet, 2013). To further explain the correlation between CDI scores and prediction error, we tested with the same method whether CDI scores correlate with the mean amplitudes of standard word responses and the mean amplitudes of the responses to familiar deviant words. For correlation analysis, the mean amplitudes were taken from time windows with significant effects in ANOVA and from electrodes with maximal amplitudes (F4 for the deviant words and C3 for the standard word).

Results

An analysis of variance (ANOVA) revealed a significant Context X Familiarity interaction [$F(1,22)=11.149$, $p=0.003$, $\eta_p^2=0.336$] in the 130-190 ms time window across age groups. Pairwise comparisons showed that only syllable [ka] completing the familiar word elicited a

larger negative mismatch response within a word than in isolation ($p=0.026$, $d=0.856$). The syllable [kɑ] also elicited a more negative mismatch response than syllable [ke] ($p=0.002$, $d=1.045$) in Context condition, whereas no difference between the stimulus types was found in No context condition (Fig. 2 and 3A).

In addition, a significant Context X Familiarity X Hemisphere interaction was found for the 40-100 ms time window [$F(1,22)=4.89$, $p=0.038$, $\eta_p^2=0.182$]. Pairwise comparisons revealed that Context condition, but not No context condition, resulted in a more positive right-hemispheric mismatch response to syllable [ke] completing the novel word compared with syllable [kɑ] ($p=0.007$, $d=0.868$). Furthermore, syllable [ke] elicited a more positive mismatch response within context than when presented in isolation ($p=0.004$, $d=0.701$, and $p=0.016$, $d=0.775$ for the left and right hemispheres, respectively), whereas responses to [kɑ] did not differ between conditions in this time window (Fig. 2 and 3A).

In later time windows, ANOVAs revealed significant effects across conditions. Syllable [ke] elicited a positive response in 12-month-olds between 190-250 ms [Group X Familiarity interaction $F(1,22)=6.837$, $p=0.016$, $\eta_p^2=0.237$] and a negative response in 24-month-olds between 310-370 ms [Group X Familiarity interaction $F(1,22)=4.53$, $p=0.045$, $\eta_p^2=0.171$] (Fig. 2).

Correlation analysis showed that at 12 months, the amplitude of the early (40-100 ms) positive right-hemispheric mismatch response to the novel word ending strongly correlated with receptive vocabulary scores ($r=0.915$, 95% confidence intervals (CI) 0.77–0.98; Fig. 3B). In contrast, no correlation was found between vocabulary scores and the mean amplitude of the mismatch response to the familiar word ending ($r=-0.199$, CI -0.76–0.43). To further determine whether the correlation between vocabulary scores and the

positive mismatch response was due to the enhancement of the response to the deviant or suppression of the response to the standard, correlation was tested also between 12-month-olds' receptive vocabulary scores and the mean amplitude of the ERP to syllable [ko] that completed the standard word in Context condition. No significant correlation was found for this comparison ($r=-0.668$, CI $-0.98-0.51$). At 24 months, no significant correlations were found between the positive right-hemispheric mismatch response and productive vocabulary scores (-0.288 , CI $-0.67-0.21$).

Discussion

The current study focused on the early phases of word learning and asked whether predictive coding supports spoken-word recognition and word learning in infants and young children. The processing of syllables as parts of words was possible in Context condition, whereas basic acoustic-phonetic and sequence-level processing of syllables was expected to take place across conditions. Significant interactions showing stronger responses to syllables completing novel and familiar words compared with the same syllables in isolation were found in 40-100 and 130-190 ms time windows, respectively. In contrast, the positive and negative mismatch responses to the acoustic-phonetic differences were observed across conditions after 200 ms at 12 and 24 months, respectively (Fig. 2, dark gray time windows). Difference in the polarity of these responses at 12 and 24 months is likely due to maturational changes (He, Hotson, & Trainor, 2009), experience with native language, or their interaction.

Compared with acoustic-phonetic effects, introducing the context syllable resulted in a striking difference in the time course and properties of responses elicited by the same syllables. Syllable [kɑ] completing the familiar word [kuk:ɑ] elicited a negative mismatch

response across groups at 130-190 ms, thus occurring earlier than acoustic-phonetic mismatch responses. This response significantly differed from the response to the same syllable in isolation, controlling for the effects of phonetic familiarity and acoustical differences between the standards and the deviants. It also significantly differed from the response to the syllable completing the novel word in Context condition. Importantly, this pattern of results also rules out the possibility that neuronal adaptation (or habituation) accounted for the findings. Although adaptation could result in smaller responses in No context than Context condition because of the more frequent activation of the same neural populations in the former, adaptation would be expected to exert a similar influence on the deviant syllables. This was not the case in Context condition, as shown by a significant difference between the responses to syllables completing familiar and novel words (Fig. 2 and 3A). Rather, the negative mismatch responses specific to the familiar word likely reflect the activation of word representations in the child brain. This interpretation is in line with the conclusions of previous studies using a similar study design in adults, associating the enhanced MMN response of the same latency (peaking ~150 ms after the syllable onset) with lexical access (Pulvermüller et al., 2001; Pulvermüller, Shtyrov, Kujala, & Näätänen, 2004; Garagnani, Shtyrov, & Pulvermüller, 2009). As proposed by Garagnani et al. (2008; Garagnani & Pulvermüller, 2011), the brain response enhancement for words could be caused by the ignition and reverberation of activity in memory circuits that have been established during word learning and that represent words in the cortex. According to the authors, these neural networks extend temporo-frontal brain areas and represent words in terms of both their perception and production via sensory-motor associations. The combined contribution of this neural circuit enhances brain responses to words, but no such circuit exists for pseudowords (Garagnani et al., 2008; Garagnani & Pulvermüller, 2011). However, it is noteworthy that according to parental reports some of our 12-month-

olds did not yet produce any words and none of them produced the word /kuk:a/. It is therefore unclear whether word representations involving aspects of speech production would be available in these children, yet their mismatch responses to the word [kuk:a] resemble MMNs to words in adults. Alternatively, our pattern of results showing enhanced and faster processing of the word-final syllable compared with the same syllable in isolation could be explained by predictive coding, which has been suggested to be available from birth (Trainor, 2012; Haden et al., 2015). According to the predictive coding hypothesis, word beginnings enable top-down lexical predictions which facilitate the encoding of auditory input and the activation of word representations (Gagnepain et al., 2012). The ignition and predictive coding accounts are not necessarily mutually exclusive, however.

Contrary to the negativity elicited for the familiar word, the final syllable of the novel word [kuk:e] elicited a positive right-hemispheric response across the age groups (Fig. 2 and 3 A). Again, acoustic-phonetic deviance or familiarity with the syllable cannot account for this positivity, because it significantly differed from the responses to the same syllable in isolation. Neither can neuronal adaptation (or habituation) account for it, since a similar degree of adaptation would be expected for both familiar and novel words, yet the responses to them had opposite polarity and differed significantly from each other. However, the pattern of results is compatible with the predictive coding hypothesis, suggesting that word beginnings create top-down predictions of familiar word endings and violations of these predictions caused by novel endings elicit lexical prediction error responses (Gagnepain et al., 2012). Given that no other reasonable explanation than predictive coding account was found for the response to the novel word [kuk:e], it is plausible that word beginnings indeed created predictions about the following speech

sounds (see Fig. 1D). These predictions generated similarly for all words have most likely affected also the processing of [kuk:ɑ], supporting the predictive coding account for the responses to this familiar word. Significant differences between the conditions are also consistent with the proposal of hierarchical predictions (Wacongne et al., 2011; Basirat et al., 2014), since the word-level predictions in Context condition likely represent a different level in the hierarchy compared with predictions in No context condition that are based solely on probabilities within the sequence.

The differences in response polarity, scalp distribution, and timing for familiar [kuk:ɑ] and novel [kuk:e] words tell us something important about the child's internal models for words. The predictive coding hypothesis proposes that responses to expected events are suppressed, whereas unexpected events elicit a prediction error. This information is projected to higher levels of the neural network for further processing and updating of the internal predictive model (Friston, 2005). Since the novel word [kuk:e] cannot be predicted, the positive mismatch response elicited by it is interpreted to reflect the prediction error, whereas the lack of such positivity to the familiar word [kuk:ɑ] is correspondingly interpreted to suggest the lack of prediction error. This implies that the word-level model may not restrict its predictions only to the word with the highest probability in the environment. Based on the match of lexical representations to the available input, the model may, at least to some degree, predict final syllables or vowels of the words included in the child's mental lexicon (Gagnepain et al., 2012; see also Cohort theory by Marslen-Wilson, 1987).

The oddball paradigm used in the present study includes extensive repetition and hence this paradigm does not fully correspond to typical word recognition situations outside the laboratory. However, it enables collecting tens of trials required for averaging the ERPs

(for other benefits, see Pulvermüller & Shtyrov, 2006). Collecting enough trials from infants would not be possible without repetition because of their limited vocabulary (e.g., their lexicons do not include tens of words with the same beginning, which was required to create and to violate predictions). In the present study design, the oddball paradigm also enabled to tease apart the effect of predictions on ERPs at sequence and word levels, as sequences of isolated syllables could be used as a control condition enabling us to extract the effect of word-level predictions. Still, one may ask whether the findings obtained with the oddball paradigm are applicable to speech listening in natural environment. According to our interpretation, the absence of a prediction error for the deviant familiar word suggests that the word-level predictive model is not based or dependent only on stimulus probabilities in the sequence. Rather, word-related brain responses likely reflect knowledge-based predictions, derived from long-term memory (Monahan & Poeppel, 2011), and the activation of long-term memory representations in the brain (Näätänen et al., 1997). Thus, the present word-level effects involving long-term memory are likely not restricted to the oddball paradigm. This interpretation is supported by previous ERP and magnetoencephalography (MEG) studies in adults, showing similar electrical or magnetic word-recognition responses when words are presented in an oddball paradigm (Pulvermüller et al., 2001, 2004) as well as when they are not presented in an oddball paradigm (MacGregor et al., 2012; for discussion, see also Pulvermüller & Shtyrov, 2006).

To further determine how our results are linked with language development in natural learning environments, we tested the hypothesis that predictive coding may be associated with the learning of new words outside the laboratory. Correlation analysis conducted between CDI vocabulary scores and mismatch response amplitudes indicated a strong correlation between receptive vocabulary scores and the strength of the prediction error at 12 months. This appears to be mainly due to the prediction error response to novel word

endings rather than response suppression for expected word endings, since the correlation involving the former was stronger and significant, whereas the correlation involving the latter was weaker and non-significant. Further research is needed to replicate these findings in larger groups of children. Nevertheless, since the correlation between vocabulary and prediction error at 12 months was quite strong, we will discuss its possible implications and explanations.

The link between word-level predictive coding and word learning at 12 months may have several accounts. Firstly, the ability of predictive coding could be determined by vocabulary size. This is, however, not compatible with the fact that no difference was found in the amplitude of the prediction error response between the age groups with considerably different vocabulary sizes (if anything, the prediction error was larger in 12-month-olds with smaller vocabulary than in 24-month-olds with larger vocabulary; see Fig. 2). Secondly, the link between vocabulary and predictive coding could be mediated by the strength of word representations: the infants with larger vocabulary could have stronger brain representations for words, which could enable them to make stronger predictions about word endings after hearing a familiar word beginning. However, if the strength of word representations accounted for the observed correlations between brain responses and vocabulary, then the strong word representations could be expected to exert the same influence on the responses to familiar words, yet no correlations were found between the brain responses to familiar words and vocabulary scores. Thirdly, the ability of word-level predictive coding could facilitate word learning. This would be reflected by a significant correlation for novel but not for familiar words, which was exactly what was found. Thus, the pattern of results is best explained by the proposal that predictive coding abilities and particularly prediction error support word learning in the early stages of language

development.

A link to natural learning of lexical items that vary in their phonological composition suggests that learning effects mediated by prediction errors do not need to be restricted to words with similar beginnings, but predictions might be formed on the basis of all context cues a child has access to, including linguistic, social (Kuhl, Tsao, & Liu, 2003; Tomasello, 2009) and visual (Fernald, Swingley, & Pinto, 2001) cues. Since CDI for receptive vocabulary measures the number and type of words that the children already understand, our findings are linked not only with the learning of phonological word forms but also with the learning of word meanings, at least indirectly. The prediction error may inform the higher-level predictive model about a need for word learning to minimize further prediction errors. It could serve as a learning signal, initiating the learning of new word forms that have been heard. The representations for word forms, being a necessary part of word-referent mapping, could then further facilitate the learning of word meanings and the establishment of full lexical representations.

It is also not excluded that the prediction error could affect learning indirectly by mediating children's cognitive resources. Since prediction is closely linked with attention in audition (Schröger, Marzecova, & SanMiguel, 2015), predictive coding could facilitate the learning of unexpected novel words by allocating attention to them for deeper processing and more efficient encoding (see Mills, Plunkett, Prat, & Schafer, 2005). Importantly, successful call of attention by strong prediction errors may result in an increase in the focus of attention to the entire learning situation, which in some cases could facilitate the mapping of to-be-learned word forms to their referents. In natural learning environments, infants with stronger predictive coding may be able to better utilize those word-learning opportunities that require fast allocation of attention, which could account for the correlation between the

amplitude of the prediction error and vocabulary size. The other side of the same coin is that children with poor predictive inference may miss some word-learning opportunities that require fast allocation of attention. To address the findings from the perspective of Schröger et al.'s (2015) framework linking prediction and attention, the context syllables used in the present study design may increase the predictability of the critical syllables by enabling top-down word-level predictions in addition to sequence-level predictions.

Predictability, in turn, modulates the gain and higher gains result in larger amplitudes of the prediction error. Large prediction errors may redirect attention to the input and modify the internal model to minimize the prediction error in the future (Friston, 2005; Schröger et al., 2015). Thus, in the present study, the early prediction error reflected as a positive mismatch response may trigger attention allocation, and the later part of the positivity may reflect focused attentive processing, which tends to be longer-lasting in younger children. Supporting this interpretation, selective attention has previously been shown to induce a similar enhancement of positive polarity in children's brain responses (Sanders, Stevens, Coch, & Neville, 2006).

Our ERP data showed a very similar pattern of prediction-related brain responses at 12 and 24 months, reflecting predictive inference at both ages. Nevertheless, no correlation was found between prediction error response and productive vocabulary at 24 months, suggesting that the reliance on prediction error as a regulator of learning may decrease as language and cognition develop. Keeping in mind the proposal that predictive coding may facilitate learning via mediating attention, the lack of correlation at 24 months might be explained by 24-month-old children's better skills in voluntary regulation of attention as compared with 12-month-olds. Consequently, the learning of new words may be independent on prediction-driven regulation of attention at 24 months.

Taken together, the present study suggests that brain responses to the same syllables markedly differ from each other across conditions: with respect to the divergence point (i.e., 20 ms from the second syllable onset), word-specific effects were early (20-170 ms) and stable across age groups, whereas phonetic effects were later (170-350 ms) and susceptible to maturation. Remarkably, our results suggest that the context syllables of novel words accelerated the detection of phonetic differences in the word-final syllables by over 100 ms, which seems to be best explained by the predictive coding hypothesis. Such processing advantage may enable more efficient encoding of novel spoken words and facilitate their learning. With respect to familiar words, previous behavioral findings have suggested that word recognition of variable, unpredicted words becomes faster between 15 and 24 months of age (Fernald, Pinto, Swingley, Weinberg, & McRoberts, 1998). Our ERP data, however, indicate that when context enables predictions, the brain processes involved in word recognition and their time course remain relatively stable during vocabulary development, supporting similar findings in later lexical-semantic processing effects as reflected by N400 (Travis et al., 2011).

Early language development has been suggested to be influenced by a number of learning mechanisms, skills, and external factors, such as statistical learning (Saffran, Aslin, & Newport, 1996; Smith & Yu, 2008), social-cognitive skills (e.g., intention-reading; Tomasello, 2009), quality of input (Cartmill et al., 2013), and social rewards (e.g., caregivers' spoken response, smile, and touch; Goldstein, King, & West, 2003). Here we propose that also predictive coding may play an important role in early vocabulary development. Predictive coding may facilitate the early learning of word forms or mediate attention allocation in young children, by means of which it may support the learning of

both word forms and the link to their referents. As a domain-general, “hard-wired” brain mechanism that is ready at birth (Näätänen et al., 2010; Trainor, 2012), predictive coding may contribute to the surprisingly early emergence of word comprehension (Bergelson & Swingley, 2013) and reflects a shared neurocognitive mechanism of learning in humans and non-human animals (Schultz et al., 1997; Näätänen et al., 2010; Steinberg et al., 2013).

References

- Basirat, A., Dehaene, S., & Dehaene-Lambertz, G. (2014). A hierarchy of cortical responses to sequence violations in three-month-old infants. *Cognition*, **132**, 137-150.
- Bergelson, E., & Swingley, D. (2012). At 6-9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, 3253-3258.
- Cartmill, E. A., Armstrong, B. F., 3rd, Gleitman, L. R., Goldin-Meadow, S., Medina, T. N., & Trueswell, J. C. (2013). Quality of early parent input predicts child vocabulary 3 years later. *Proceedings of the National Academy of Sciences of the United States of America*, **110**, 11278-11283.
- Conboy, B. T., & Kuhl, P. K. (2011). Impact of second-language experience in infancy: Brain measures of first- and second-language speech perception. *Developmental Science*, **14**, 242-248.
- Fenson, L., Dale, P. S., Reznick, J. S., Bates E., Thal, D. J., & Pethick, S. J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, Ser. no. 242, vol. 59.

- Fernald, A., Pinto, J. P., Swingle, D., Weinberg, A., & McRoberts, G. W. (1998). Rapid gains in speed of verbal processing of infants in the 2nd year. *Psychological Science*, **9**, 228-231.
- Fernald, A., Swingle, D., & Pinto, J. P. (2001). When half a word is enough: Infants can recognize spoken words using partial phonetic information. *Child Development*, **72**, 1003-1015.
- Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **360**, 815-836.
- Friston, K. (2009). The free-energy principle: A rough guide to the brain? *Trends in Cognitive Sciences*, **13**, 293-301.
- Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, **11**, 127-138.
- Gagnepain, P., Henson, R. N., & Davis, M. H. (2012). Temporal predictive codes for spoken words in auditory cortex. *Current Biology*, **22**, 615-621.
- Garagnani, M., & Pulvermüller, F. (2011). From sounds to words: A neurocomputational model of adaptation, inhibition and memory processes in auditory change detection. *Neuroimage*, **54**, 170-181.
- Garagnani, M., Shtyrov, Y., & Pulvermüller, F. (2009). Effects of attention on what is known and what is not: MEG evidence for functionally discrete memory circuits. *Frontiers in Human Neuroscience*, **3**, 10.
- Garagnani, M., Wennekers, T., & Pulvermüller, F. (2008). A neuroanatomically grounded Hebbian-learning model of attention-language interactions in the human brain. *European Journal of Neuroscience*, **27**, 492-513.

- Goldstein, M. H., King, A. P., & West, M. J. (2003). Social interaction shapes babbling: Testing parallels between birdsong and speech. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 8030-8035.
- Haden, G. P., Nemeth, R., Torok, M., & Winkler, I. (2015). Predictive processing of pitch trends in newborn infants. *Brain Research*, **1626**, 14-20.
- He, C., Hotson, L., & Trainor, L. J. (2009). Maturation of cortical mismatch responses to occasional pitch change in early infancy: Effects of presentation rate and magnitude of change. *Neuropsychologia*, **47**, 218-229.
- Kuhl, P. K., Tsao, F. M., & Liu, H. M. (2003). Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 9096-9101.
- Lyytinen P. (1999). *Communicative and language development inventory*. Jyväskylä (Finland): Niilo Mäki Institute.
- MacGregor, L. J., Pulvermüller, F., van Casteren, M., & Shtyrov, Y. (2012). Ultra-rapid access to words in the brain. *Nature Communications*, **3**, 711.
- Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, **25**, 71-102.
- Mills, D. L., Plunkett, K., Prat, C., & Schafer, G. (2005). Watching the infant brain learn words: Effects of vocabulary size and experience. *Cognitive Development*, **20**, 19-31.
- Näätänen, R., Astikainen, P., Ruusuvirta, T., & Huotilainen, M. (2010). Automatic auditory intelligence: An expression of the sensory-cognitive core of cognitive processes. *Brain Research Reviews*, **64**, 123-136.

- Näätänen, R., Gaillard, A. W., & Mäntysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta Psychologica*, **42**, 313-329.
- Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., Vainio, M., Alku, P., Ilmoniemi, R.J., Luuk, A., Allik, J., Sinkkonen, J., & Alho, K. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature*, **385**, 432-434.
- Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology*, **118**, 2544-2590.
- Pernet, C. R., Wilcox, R., & Rousselet, G. A. (2013). Robust correlation analyses: False positive and power validation using a new open source Matlab toolbox. *Frontiers in Psychology*, **3**, 606.
- Poeppel, D., & Monahan, P. J., (2011). Feedforward and feedback in speech perception: Revisiting analysis by synthesis. *Language and Cognitive Processes*, **26**, 935-951.
- Pouget, A., Beck, J. M., Ma, W. J., & Latham, P. E. (2013). Probabilistic brains: Knowns and unknowns. *Nature Neuroscience*, **16**, 1170-1178.
- Pulvermüller, F. (1999). Words in the brain's language. *Behavioral and Brain Sciences*, **22**, 253-279.
- Pulvermüller, F., Kujala, T., Shtyrov, Y., Simola, J., Tiitinen, H., Alku, P. et al. (2001). Memory traces for words as revealed by the mismatch negativity. *Neuroimage*, **14**, 607-616.
- Pulvermüller, F., & Shtyrov, Y. (2006). Language outside the focus of attention: The mismatch negativity as a tool for studying higher cognitive processes. *Progress in*

Neurobiology, **79**, 49-71.

Pulvermüller, F., Shtyrov, Y., Kujala, T., & Näätänen, R. (2004). Word-specific cortical activity as revealed by the mismatch negativity. *Psychophysiology*, **41**, 106-112.

Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, **2**, 79-87.

Rousseeuw, P. J. (1984). Least median of squares regression. *Journal of the American Statistical Association*, **79**, 971–880.

Rousseeuw, P. J., & Van Drissen, K. (1999). A fast algorithm for the minimum covariance determinant estimator. *Technometrics*, **41**, 212–223.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, **274**, 1926-1928.

Sanders, L. D., Stevens, C., Coch, D., & Neville, H. J. (2006). Selective auditory attention in 3- to 5-year-old children: An event-related potential study. *Neuropsychologia*, **44**, 2126-2138.

Schröger, E., Marzecova, A., & SanMiguel, I. (2015). Attention and prediction in human audition: A lesson from cognitive psychophysiology. *The European Journal of Neuroscience*, **41**, 641-664.

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, **275**, 1593-1599.

Smith, L.B., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, **106**, 333-338.

Steinberg, E. E., Keiflin, R., Boivin, J. R., Witten, I. B., Deisseroth, K., & Janak, P. H.

(2013). A causal link between prediction errors, dopamine neurons and learning.

Nature Neuroscience, **16**, 966-973.

Tomasello, M. (2009). The usage-based theory of language acquisition. In E. Bavin (Ed.),

Handbook of child language (pp. 69-87). Cambridge: Cambridge University Press.

Trainor, L. J. (2012). Predictive information processing is a fundamental learning

mechanism present in early development: Evidence from infants. *International Journal of Psychophysiology*, **83**, 256-258.

Travis, K. E., Leonard, M. K., Brown, T. T., Hagler, D. J., Jr, Curran, M., Dale, A. M. et al.

(2011). Spatiotemporal neural dynamics of word understanding in 12- to 18-month-old-infants. *Cerebral Cortex*, **21**, 1832-1839.

Verboten, S., & Hubert, M. (2005). LIBRA: a MATLAB library for robust analysis.

Chemometrics and Intelligent Laboratory Systems, **75**, 127–136.

Wacongne, C., Labyt, E., van Wassenhove, V., Bekinschtein, T., Naccache, L., &

Dehaene, S. (2011). Evidence for a hierarchy of predictions and prediction errors in human cortex. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 20754-20759.

Winkler, I. (2007). Interpreting the mismatch negativity. *Journal of Psychophysiology*, **21**,

147-163.

Ylinen, S., Huuskonen, M., Mikkola, K., Saure, E., Sinkkonen, T., & Paavilainen, P.

(submitted). Predictive coding of phonological rules in auditory cortex: A mismatch negativity study.

A Stimulation in No context condition

B Stimulation in Context condition

C Predictions in No context condition

D Predictions in Context condition

Fig. 1. Excerpts of stimulus sequences in the two conditions. The same syllables of interest are introduced in (A) no context condition; (B) context condition, where they form a

familiar word (blue) or a novel word (magenta) with the context (gray). Hypothesized predictions (marked with arrows), their hypothesized comparisons with auditory input and consequent brain responses hypothesized to be generated by the stimuli in (C) No context and (D) Context conditions. There were no specific hypotheses for two standards following deviants (they were omitted from analysis). Note the two levels of prediction in Context condition (D). Familiar word beginnings may predict the endings of both familiar words /kuk:o/ and /kuk:a/, but likely the prediction for /ko/ is stronger (marked with bold) due to interaction with the sequence-level predictions. S denotes standard, D deviant, and MMR mismatch response.

Fig. 2. Brain responses at 12 and 24 months. (A) Event-related potentials to the syllables of interest, presented among context syllables (Context condition) and in isolation (No context condition) at a right frontal electrode site. (B) Deviant-minus-standard difference waveforms at left and right frontal electrode sites. In both (A) and (B), time windows with significant word-level and phonetic effects have been marked with light gray and dark gray, respectively, and arrows point to those deviant responses that are significantly different compared with all other deviant responses across conditions in that time window.

Fig. 3. Word-level prediction effects, dependent on context. (A) The difference between response amplitudes in the two conditions (context minus no context $\pm$ standard error of the mean) in 40-100 and 130-190 ms time windows, averaged across age groups (N=24) and electrode sites. (B) The correlation between communicative development inventory (CDI) score for receptive vocabulary and the mean amplitudes of early positive prediction-error responses to novel words in 12-month-olds (N=10) at a right frontal electrode site.

Table 1. Hypothesized brain responses to different violations of predictions for the two deviant syllables in No context and Context conditions. Note that the comparison processes of auditory input with predictions is hypothesized to result in different effects between the conditions and between deviant syllables in Context condition.

	Deviant syllable [kɑ]	Deviant syllable [ke]
No context condition	Mismatch with sequence-level predictions: MMR	Mismatch with sequence-level predictions: MMR
Context condition	Mismatch with sequence-level predictions, match with word-level predictions: Lexical activation	Mismatch with sequence-level predictions, mismatch with word-level predictions: Prediction error