Seven-year-olds' recollection of non-adjacent dependencies after overnight retention

Gesa Schaadt [1, 2,] *, Mariella Paul [1, 3]*, R. Muralikrishnan [4], Claudia Männel [1, 2], Angela D. Friederici [1]

[1] Department of Neuropsychology, Max Planck Institute for Human Cognitive and Brain Sciences, Stephanstr. 1a, 04103 Leipzig, Germany

[2] Department of Neurology, Max Planck Institute for Human Cognitive and Brain Sciences, Stephanstr. 1a, 04103 Leipzig and Clinic for Cognitive Neurology, Medical Faculty, Leipzig University, Liebigstr. 16, 04103 Leipzig, Germany

[3] Berlin School of Mind and Brain, Humboldt-Universität zu Berlin, Luisenstraße 56, 10117 Berlin, Germany

[4] Max Planck Institute for Empirical Aesthetics, Grüneburgweg 14, 60322 Frankfurt am Main, Germany

*These authors contributed equally

Correspondence concerning this article should be addressed to Gesa Schaadt, Department of Neuropsychology, Max Planck Institute for Human Cognitive and Brain Sciences, Stephanstr. 1a, 04103 Leipzig, Germany, Tel: +49 341 9940-2465, Email: Schaadt@cbs.mpg.de

Highlights

- 7-year-old's non-adjacent dependency learning in a foreign language tested
- Children gave grammaticality judgments while electroencephalography was recorded
- Brain responses revealed children's learning of non-adjacent dependencies
- Brain responses after overnight retention showed different polarity
- Children recollected dependencies after sleep associated with representation change

Abstract

Becoming a successful speaker depends on acquiring and learning grammatical dependencies between neighboring and non-neighboring linguistic elements (non-adjacent dependencies; NADs). Previous studies have demonstrated children's and adults' ability to distinguish NADs from NAD violations right after familiarization. However, demonstrating NAD-recollection and processing after retention is crucial to demonstrate NAD-learning. We tested 7-year-old's NAD-learning in a natural, non-native language on one day and NAD-recollection on the next day by means of event-related potentials (ERPs). Our results revealed ERPs with a more positive amplitude to NAD violations than correct NADs after familiarization on day one, but ERPs with a more negative amplitude to NAD violations on day two. This change from more positive to more negative ERPs to NAD violations possibly indicates that children's representations of NADs changed during an overnight retention period, potentially associated with children's NAD-learning. Indeed, both ERP patterns (i.e., day one: positive, day two: negative) were related to stronger behavioral improvement (i.e., more correct answers on day two compared to day one) in a grammaticality judgment task from day one to day two. We suggest these findings to indicate that children showing more correct answers at day two, compared to day one, successfully built associative representations of NADs on day one and then successfully strengthened these associations during overnight retention, revealing NAD-recollection on day two. The present results suggest that 7-year-olds readily track NADs in a natural, non-native language and are able to show recollection after a retention period involving sleep, providing strong evidence of NAD-recollection.

Keywords. non-adjacent dependencies, ERPs, recollection, children, development

1. Introduction

Language is made up of different building blocks, combined together to form sentences. The grammar of a given language defines the rules for these combinations. For example, grammatical rules define that determiners can be combined with nouns (*The girl*), but not with verbs (**The give*). Grammatical dependencies can be formed not only between neighboring elements, but also between non-neighboring elements of a sentence. For example, in the sentence *The girl*$_{Sg}$ *smiles*$_{g}$, *girl*$_{Sg}$ *and -s*$_{Sg}$ form a grammatical dependency (i.e., number agreement) that spans one element (smile). In theory, these dependencies can span an arbitrary number of elements, as demonstrated in the following example: *The girl*$_{Sg}$ *who visited us yesterday smiles*$_{Sg}$. This type of dependencies, called non-adjacent dependencies (NADs), are important grammatical rules of language, such that becoming a proficient speaker highly depends on acquiring these rules.

Adults have been shown to be able to process and learn NADs in a number of behavioral studies (e.g. Frost & Monaghan, 2016; Gómez, 2002; Newport & Aslin, 2004; Peña, Bonatti, Nespor, & Mehler, 2002). For example, Gómez (2002) exposed adults to an artificial language containing NADs in the form of three-syllable strings. In artificial language learning tasks, NADs are often realized as AXC structures, with A and C being the dependent elements and X being variable elements. After familiarization to these strings, participants were shown a mixture of strings, either containing familiarized NADs or NAD violations; they were asked to indicate whether a given string followed the rules of the familiarized artificial language. The results showed that adults are in principle able to learn NADs (Gómez, 2002). However, adults' NAD-learning is somewhat restricted, as it has been shown that adults only successfully learn NADs when phonological cues between dependent elements were provided (Mueller, Bahlmann, & Friederici, 2008 ; Newport and Aslin, 2004; Peña et al., 2002). Taken together, behavioral studies demonstrated that adults are able to learn NADs in an artificial language. Becoming a successful speaker, however, already starts in early infancy; and even infants have been shown to be able to learn NADs (Gómez and Maye, 2005; Lany & Gómez, 2008), but with some restrictions depending on how the NADs are presented (e.g. Höhle, Schmitz, Santelmann, & Weissenborn, 2006; Santelmann & Jusczyk, 1998). For example, Gómez and Maye (2005) exposed infants to AXC grammars using the Head Turn Preference Procedure (Nelson et al., 1995), which measures children's looking time towards an auditory stream played on either side of the child. Gómez and Maye (2005) first familiarized children with the AXC grammar (i.e., NADs), which was then followed by the presentation of correct or incorrect (i.e., containing a violation) NADs via loudspeakers, either located on the left or the right side of the child. Fifteen-month-old infants oriented more towards the familiarized stimuli (i.e., correct NADs) than to violations (i.e., incorrect NADs), indicating that 15-month-olds learned the AXC grammar. Taken together, both adults and children are able to learn NADs in principle. However, the processes underlying NAD-learning cannot be fully understood by using behavioral methods alone, but should be supplemented by other methods, such as measuring event-related potentials (ERPs).

Mueller, Oberecker, and Friederici (2009) used ERPs to investigate the learning of NADs that were embedded in natural speech in a foreign language (Italian). During familiarization, they exposed German native speakers with no prior knowledge of Italian, to Italian sentences containing NADs (e.g. "*La sorella sta cantando*"; *the sister is singing*). In testing phases, participants heard a mixture of correct sentences and incorrect sentences containing NAD violations (e.g. "*La sorella sta cantare*"; *the sister is sing∅*). By comparing ERPs to incorrect sentences with ERPs to correct sentences during testing phases, a series of studies could show that both infants (under passive listening conditions, i.e. without a task; Friederici, Mueller, & Oberecker, 2011) and adults (under active conditions, i.e. with a task; Mueller et al., 2009) are able to learn these NADs embedded in a miniature version of Italian. Infants showed a more positive ERP response to incorrect compared to correct NADs, while adults showed a more negative ERP response. Thus, infants and adults might use different learning mechanisms and develop different representations of the NADs. Specifically, it has been suggested that infants learn NADs more automatically than adults do, also reflected in infants' ability to learn under passive listening, which adults struggle to do (Mueller, Friederici, & Männel, 2012). Interestingly, Mueller, Friederici, and Männel (2018) showed that children up to the age of 2 years are able to learn NADs under passive listening conditions, while 4-year-olds, similar to adults (Mueller et al., 2012) struggle to do so and may need active task conditions. It has been suggested that the specific need for an active task is associated with a switch in learning mechanisms from associative, bottom-up learning (allowing learning under passive listening conditions) to controlled, top-down learning (hindering learning under passive listening conditions, but facilitating learning under active task conditions; see Skeide & Friederici, 2016). This switch is likely associated with prefrontal cortex (PFC) maturation (Skeide & Friederici, 2016), which reaches near adult-like maturity around the age of 7 years (Huttenlocher, 1990). Taken together, NAD-learning mechanisms change during development, which is possibly linked to PFC development.

Although previous studies (Friederici et al., 2011; Mueller et al., 2009; 2012) convincingly demonstrated that individuals can differentiate familiarized NADs from NAD violations, NAD-learning was always tested on the day of the familiarization itself. This testing procedure likely provides a measure of whether participants have formed a representation of the familiarized items, which can then be compared to test items (similarity-based learning; see Opitz & Hofmann, 2015). Whether this should be interpreted as NAD-learning, however, is a matter of discussion, because the knowledge of the underlying rules that characterize the (artificial) grammar only builds up over time (Opitz & Hoffman, 2015) and might not be fully present immediately after a relatively brief familiarization. In order to investigate whether participants have learned the underlying rules, several studies have investigated recollection of grammatical rules after a retention period. For example, Fischer, Drosopoulos, Tsen, and Born (2006) investigated the effect of a retention period on artificial grammar learning in adults. The authors showed that before sleep, adults did not perform above chance in a generation task, during which participants had to predict the next letter in a string based on the artificial grammar. However, after a retention period involving sleep participants could solve the task successfully, which was not the case after a retention period without sleep. Not only adults, but especially infants and children were shown to benefit from a retention period (particularly when retention involved sleep) for artificial language learning (Backhaus, Hoeckesfeld, Born, Hohagen, & Junghanns, 2008; Hupbach, Gómez, Bootzin, & Nadel, 2009) and for generalizing learned information to new input (Gómez, Bootzin, & Nadel, 2006). A study by Friedrich, Wilhelm, Mölle, Born, and Friederici (2017) showed that representations of learned associations change during the course of a retention period, offering insight into how learned associations can be recollected after a retention period. In this study, infants were exposed to object-word pairs followed by a retention period that either involved a long nap, a short nap, or no sleep. Before retention, there was no evidence for learning of the object-word pairs and neither did the group without sleep show any sign of learning after retention. In contrast, infants who had a short retention period involving sleep showed consolidation of the object-word pairs. However, the ERPs only revealed a phonological association between the word and object, but no evidence for a lexical-semantic representation of the object-word pairs in long-term memory. Only those children who had a longer consolidation period involving sleep also showed neurophysiological evidence of lexical-semantic representations of word meaning in long-term memory (Friedrich et al., 2017). Thus, this study demonstrates that children benefit from a retention period involving sleep, which most likely leads to the ERP effects of successful recollection after the retention period. Given these promising findings showing a beneficial effect of a retention period involving sleep on long-term memory consolidation, we aimed at investigating the effect of retention involving sleep on the recollection of NADs as important grammatical rules of language.

Thus, in the present ERP study, we investigated 7-year-old children's recollection of NADs embedded in a miniature version of a foreign language (i.e., Italian), using the same paradigm as Mueller and colleagues (2009), including a grammaticality judgment task. We invited our participants on two consecutive days, ensuring a retention period involving sleep to test recollection. If we can show recollection of NADs on day two, we provide evidence that children learned the NADs, which goes beyond showing processing differences between correct and incorrect NADs on the same day when familiarization took place. We tested 7-year-olds because they have been shown to be able to successfully perform offline behavioral tasks assessing artificial grammar learning (Raviv & Arnon, 2018; Shufaniya & Arnon, 2018), most likely associated with 7-year-old's advanced PFC maturation (Huttenlocher, 1990), playing a crucial role in NAD-learning (Friederici, Mueller, Sehm, and Ragert, 2013). According to the procedure of Mueller and colleagues (2009) in adults, children listened to only correct stimuli (i.e., Italian sentences) during the four learning phases on the first testing day. Each learning phase was followed by a testing phase, during which children listened to incorrect stimuli containing NAD violations intermixed with correct stimuli following the familiarized NAD rule. During the testing phases, children were required to behaviorally indicate whether or not a given stimulus belonged to the language they were familiarized with in the learning phases (i.e., grammaticality judgment task). On the following day, we tested recollection by asking children to perform only the four testing phases, again including the grammaticality judgment task. To capture consolidation and recollection of NAD-learning on the next day, we specifically focused on the change in behavior from day one to day two. Successful recollection will be reflected in behavioral improvement from day one to day two (i.e., more correct grammaticality judgments on day two compared to day one). If learning takes place, we expect that children's ERP responses are associated with their improvement in the number of correct grammaticality judgments from day one to day two.

2. Materials and Methods

2.1 Participants

For the present experiment, 49 children were invited. The datasets of 36 children (20 boys) with a mean age of 7.22 years [*Standard Deviation (SD)* = 0.36] entered the final analyses (i.e., the datasets of 13 children were excluded due to movement and perspiration artifacts in the EEG). Children visited the first and second school grade. All participants were German monolinguals and none of the children had any known hearing deficits or neurological problems. In order to ensure that the Italian sentences used for the present study were foreign to the children and thus, functioned as an “artificial” language, we asked the parents about the child’s experience with foreign languages and specifically with the Italian language. One of the 36 children visited a bilingual French-German kindergarten and at school, 11 of the 36 children learned a second language, with two children learning French and nine children learning English. Thus, none of the children had any specific experience with the Italian or Spanish (Spanish and Italian consist of the same NADs) language.

The study followed American Psychological Association (APA) standards in accordance with the declaration of Helsinki from 1964 (World Medical Association, 2013) and was approved by the ethics committee of the University Leipzig. Parental written consent was obtained after children and parents had been informed about the procedure and agreed to participation.

2.2 Stimulus material

Mueller and colleagues (2009) provided the stimuli for the present study. They consisted of simple Italian sentences, containing an NAD between an auxiliary and a main verb’s suffix. Sentences were made up of one of two noun phrases (*il fratello*, the brother; *la sorella*, the sister), one of two auxiliaries (*può*, to be able to, first person singular; *sta*, to be, first person singular), and one of 32 verbs. Verbs could either occur in infinitive (e.g., *arrivare*) or in gerund form (e.g., *arrivando*). Between the auxiliary and the verb suffix was a non-adjacent grammatical dependency, such that the auxiliary *sta* required the gerund form -*ando* and the auxiliary *può* required the infinitive form -*are*. In total, 128 correct sentences were generated. All correct sentences were spoken by a female native Italian speaker and digitally recorded. Subsequently, the auditory material was segmented and normalized using the ReZound software. Incorrect sentences were produced by combining auxiliaries with the incorrectly suffixed verbs from a different, correct sentence. This was done by a cross-splicing procedure at the beginning of each verb. In each sentence, the verb was thus exchanged with a verb from a different sentence. To control for splicing effects across conditions, correct sentences were spliced in the same manner.

2.3 Experimental Procedure

Participants were invited for two consecutive days. On the first testing day, participating children and their parents were verbally informed about the procedure. Children were asked to provide consent to participate and parents gave written informed consent on behalf of their children. Our experiment on the first day comprised four alternating learning and testing phases. In each learning phase, participants were presented with 64 correct sentences (256 in total across all learning phases). After a learning phase, a testing phase followed where participants were presented with correct sentences and incorrect versions of the sentences containing NAD violations. Each testing phase consisted of 8 correct and 8 incorrect sentences (64 sentences across all four testing phases). Please note that each testing phase contained different auxiliary-verb-suffix-triplets compared to the preceding learning phase to ensure that participants learned NADs (and not auxiliary-verb-suffix triplets).

For the ERP experiment, participants sat in a sound-attenuated booth in front of a computer screen and stimuli were presented via loudspeaker using Presentation® software Version 14 (Neurobehavioral Systems Inc., Berkely CA, USA). Children were instructed by using a cover story, where they were asked to support an adventurer, needing help to decide whether the sentences in a foreign language are correct or incorrect. Further, they were told that it is important to listen carefully, because otherwise the adventurer would not be able to continue his journey around the world. After the instruction, the experiment started with the first learning phase, in which participants passively listened to the correct NAD sentences. A fixation cross was continuously presented in the middle of the screen to reduce extreme eye-movements. Sentences were presented in a pseudo-randomized order, such that each sentence beginning (i.e., *il fratello sta, la sorella sta, il fratello può, la sorella può*) was not presented more than three-times in a row; and such that a verb could only be repeated every third sentence. From the beginning of each sentence to the beginning of the following sentence, there was an inter-stimulus-interval (ISI) of 3000 ms. Each learning phase (in total 4 learning phases) was followed by a grammaticality judgment task to test for learning effects, with pseudo-randomized presentation following the above mentioned criteria for pseudo-randomization and pseudo-randomized presentation of incorrect and correct sentences, such that correct or incorrect sentences could only be presented twice in a row. Children were required to give grammaticality judgments on each stimulus (i.e., correct vs. incorrect) by using a button-press response device. The trials started with a fixation cross that was presented for 1000 ms, before one of the correct or incorrect sentences was presented. After an ISI of 3000 ms, the simultaneous display of a happy (indicating correct) and a sad face (indicating incorrect) prompted participants to judge the grammatical correctness of the sentence via the provided response keys. The response key assignment (right / left) to the answer type (correct / incorrect) during the testing phases was counterbalanced across participants. Each learning phase lasted for about 3.5 min and each testing phase lasted for about 10 min (i.e., depending on the child's response times), summing up to a total experimental time of around 60 min.

For the second testing day, our experiment only comprised four testing phases (i.e., grammaticality judgment task as described above), but no learning phases, to investigate NAD-learning after a retention period including sleep. Stimuli used for the second testing day were not identical to those used on the first testing day. As participants were not presented with the learning phases on the second testing day, the total experimental time was reduced (i.e., around 40 min). EEG was recorded during the whole experiment on both testing days. Behavioral data (i.e., error rates; response times) were recorded for each participant during testing phases on testing day one and testing day two.

2.4 EEG Recordings and Analysis

Continuous EEG was recorded with an EGI (Electrical Geodesics, 1998) 128-electrode array (see Figures 1, 2, and 3 for schematic illustration). The vertex (recording site Cz) was chosen as online reference. For the EGI high input impedance amplifier, impedances were kept below 75 kΩ. The sampling rate was 500 Hz and all channels were pre-processed on-line by means of 0.01 – 200-Hz band-pass filter. In addition, vertical and horizontal eye movements were monitored with a subset of the 128 electrodes.

For off-line EEG analysis, we used the Fieldtrip toolbox for EEG/MEG analysis (Oostenveld, Fries, Maris, & Schoffelen, 2011) and the MATLAB® version R2017b (The MathWorks Inc., 2017). Before preprocessing, EEG data was manually scanned for electrodes with bad or missing signal. Those electrodes were excluded from the respective data set. Note, however that the number of excluded electrodes never exceeded 6 out of 128 (i.e., < 5%) and that excluded electrodes differed across participants. Thereafter, data were off-line re-referenced to the average of all EEG electrodes. Before data were filtered, the sampling rate was reduced to 250 Hz. We then applied a digital low-pass filter of 30 Hz (Kaiser-windowed finite-impulse response high-pass filter, half-amplitude cutoff (-6 dB) of 0.3 Hz, transition width of 0.3 Hz) to remove muscle artifacts and a high-pass filter of 0.3 Hz (Kaiser-windowed finite-impulse response low-pass filter, half-amplitude cutoff (-6 dB) of 30 Hz, transition width of 5 Hz), to remove very slow drifts. In a next step, we extracted trials of –200 to 2000 ms time-locked to the onset of the critical verb (i.e., containing either the correct or incorrect suffix). Across all remaining trials, we identified muscle artifacts with a distribution-based identification approach. We set the rejection threshold to z = 7.0. Trials were visually scanned and, if applicable, further trials with severe artifacts were manually marked and removed. To remove eye-movement artifacts, an independent-component analysis (ICA; Makeig, Bell, Jung, & Sejnowski, 1996) was performed. ICA components were visually scanned and eye movement-related components removed. Before individual averages were computed (baseline corrected from –200 to 0 ms relative to verb-onset), the removed electrodes with bad or missing signal were interpolated by using spherical spline interpolation (Perrin, Pernier, Bertrand, & Echallier, 1989). In a second step, grand averages were computed in relation to the suffix onset for the learning phases (separately for the first and second halves of the experiment) and for the testing phases at day one and day two, separately for verbs containing correct suffixes (i.e., NAD was not violated) and verbs with incorrect suffixes (i.e., NAD was violated).

2.5 Statistical Analysis

For statistical analyses, we used the Statistical Package for the Social Sciences (SPSS) Software Version 24 (IBM; Walldorf, Germany).

2.5.1 Behavioral data

For each testing day, statistical means of response times (RTs) in ms and correct answers in percent were calculated for each participant. To analyze whether RTs and correct answers differed between day one and day two, we calculated dependent *t*-Tests. In a next step, we performed binomial tests for each child to determine whether performance (i.e., correct answers in percent) was above chance level in the grammaticality judgment NAD task. Finally, we obtained a score indicating whether children's task performance changed (i.e., number of correctly answered trials) from day one to day two by calculating the difference between correct answers at day two and the correct answers at day one.

2.5.2 EEG data

To statistically analyze the ERP data, we defined two frontal regions of interest (ROIs), two central ROIs, and two parietal ROIs for each hemisphere (i.e., left and right; see Figures 1, 2, 3 and Luu & Ferree, 2005). Further, we defined ROIs for the midline (see Figures 1, 2, 3 and Luu & Ferree, 2005). ERP analyses were performed on six time windows (TW) of 200 ms each. The suffix-onset (-*are* and -*ando*) served as criterion for TW definition, as it is the earliest point, at which a correct sentence can be distinguished from an incorrect sentence. On average, suffix onset occurred at 267 ms (range: 138 – 408 ms) relative to the onset of the verb stem, such that we defined the first TW of interest to start 300 ms after verb onset.

To identify significant ERP effects of learning across the experiment at day one, we contrasted ERPs in response to the critical suffixes during the first and second halves of the experiment. In order to do so, we calculated a three-factorial repeated measures analyses of variance (ANOVAs) with the within-subject factors condition (first half, second half), region (left frontal, centro frontal, right frontal, left central, centro central, right central, left parietal, centro parietal, right parietal), and TW (300–500 ms, 500–700 ms, 700–900 ms, 900–1100 ms, 1100–1300 ms, 1300–1500 ms). If effects involving the factor condition reached significance ($p < .05$), post-hoc pairwise comparisons were computed and P-Values were Bonferroni-corrected.

To identify significant ERP differences between the processing of correct and incorrect suffixes during testing phases, we calculated a three-factorial ANOVA with the within-subject factors condition (correct, incorrect), region (left frontal, centro frontal, right frontal, left central, centro central, right central, left parietal, centro parietal, right parietal), and TW (300–500 ms, 500–700 ms, 700–900 ms, 900–1100 ms, 1100–1300 ms, 1300–1500 ms) separately for the first and second testing day. If effects involving the factor condition reached significance ($p < .05$), post-hoc pairwise comparisons were computed and P-Values were Bonferroni-corrected.

In a further step, we analyzed whether significant ERP effects could predict the change in task performance from day one to day two. In order to do so, we calculated the ERP difference waves between condition levels of those contrasts, for which the above-described ANOVAs revealed statistically significant effects. We then calculated a linear regression analysis with the ERP difference waves as predictors and the change in task performance from day one to day two (correct answers day two – correct answers day one) as dependent variable. Backward entry was chosen as regression method with $p \leq .10$ as criterion for predictors to be entered in the final model.

3. Results

3.1 Behavioral Results

Correct answers in percent at day one (*mean* = 48.65%; *SD* = 5.90) did not differ significantly from correct answers at day two (*mean* = 50.73%; *SD* = 6.18; t (35) = -1.49; $p = .15$). RTs were significantly shorter on day two (*mean* = 7489.13 ms; *SD* = 3843.89) compared to day one (*mean* = 11552.69 ms; *SD* = 4847.99; t (35) = 5.77; $p < .001$). The binomial test showed that 58.2% correct answers in percent indicated above chance-level performance. When using this criterion, we could identify three children performing above chance level on day one and three children on day two (only partially overlapping). Further, the *mean* change in behavior was 2.09% (*SD* = 8.41; min = -15.65; max = 17.15), indicating that some children showed more correct answers at day two compared to day one (i.e., positive values) and some children showed less correct answers at day two compared to day one (i.e., negative values).

3.2 EEG Results

3.2.1 Learning phases day one

We did not find any significant effects involving the factor condition (see Figure 1).

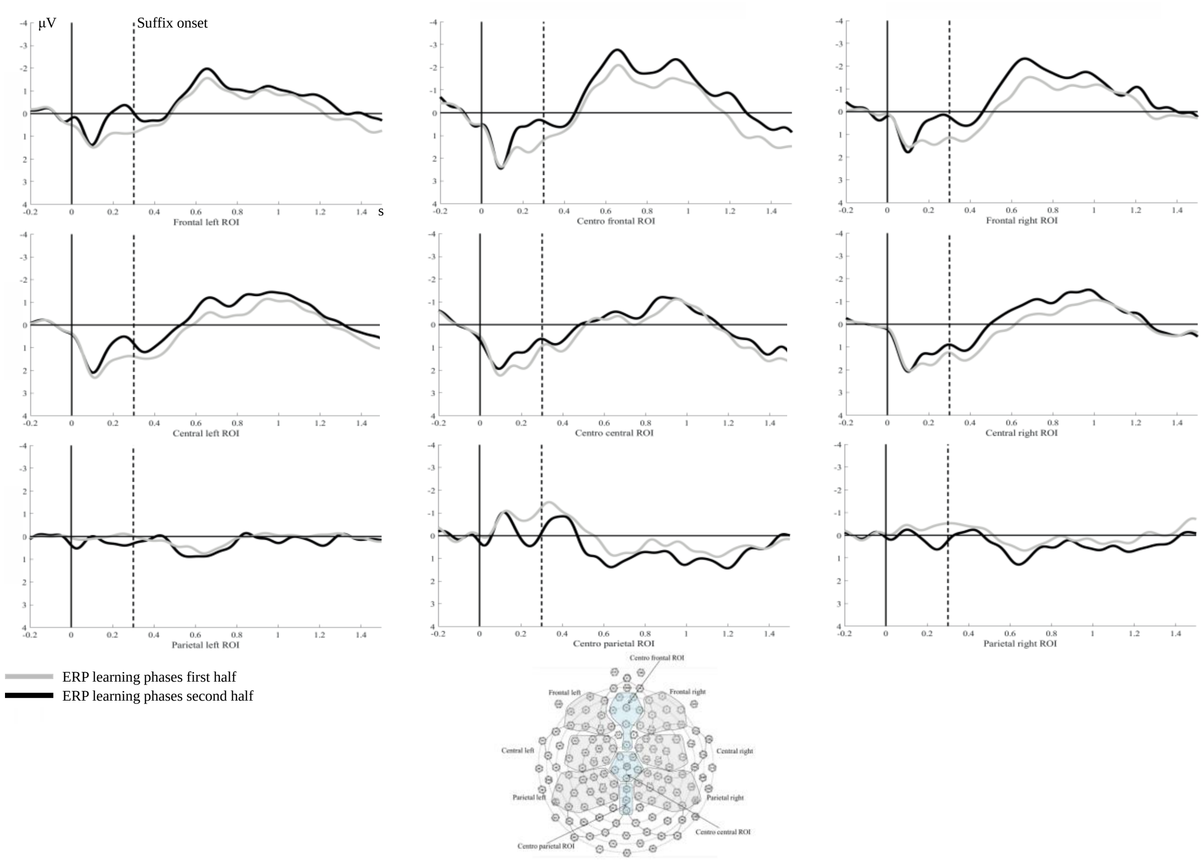

Figure 1. Event-related potentials (ERPs) of the learning phases at day 1. Illustrated are the mean ERPs in response to the correct suffixes during the first half of the learning phases (grey line) and to the correct suffixes during the second half of the learning phases (black line) averaged for left, middle, and right frontal, central, and parietal regions of interest (ROIs; see schematic head for details on electrodes).

3.2.2 Testing phases day one

We found a significant interaction between the factors condition, region, and TW [F (40, 1400) = 1.73; p = .04; η^2 = .06], which could be explained by a more positive ERP response to incorrect suffixes compared to correct suffixes for the TW 1100–1300 ms at the left frontal region (p = .01) and at the centro frontal region (p = .02) (see Figure 2). Thus, we found a more positive ERP response to incorrect compared to correct suffixes between 1100 and 1300 ms, that is, between 800 and 1000 ms after suffix onset.

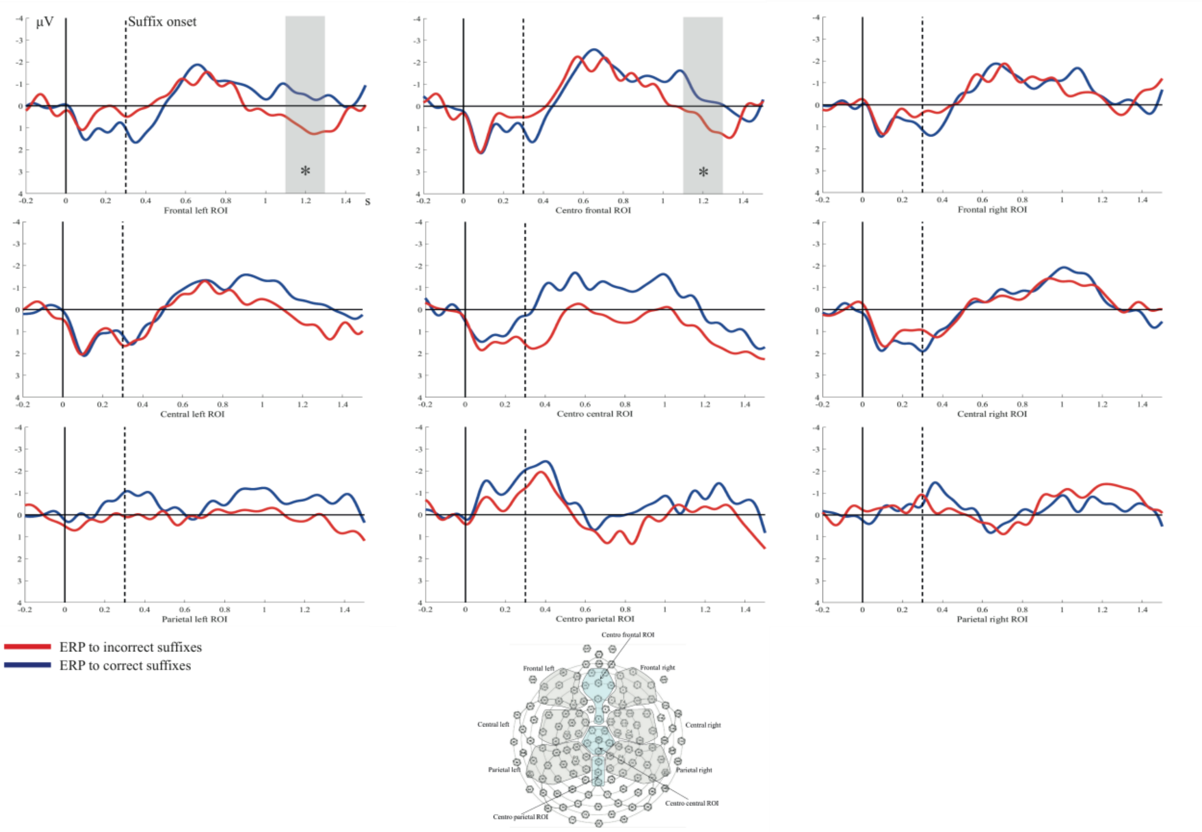

Figure 2. Event-related potentials (ERPs) of the testing phases at day 1. Illustrated are the mean ERPs in response to the correct suffixes containing the nonadjacent dependency (blue line) and to the incorrect suffixes violating the nonadjacent dependency rule (red line) averaged for left, middle, and right frontal, central, and parietal regions of interest (ROIs; see schematic head for details on electrodes). Grey bars and asterisk indicate time windows and regions with significant differences between the ERPs of the two conditions (* p < .05).

3.2.3 Testing phases day two

We found a significant interaction between the factors condition, region, and TW [F (40, 1400) = 1.61; p = .05; η^2 = .05], which could be explained by a more negative ERP response to incorrect suffixes compared to correct suffixes for the TW 900–1100 ms at the left frontal region (p = .05) and at the left central region (p = .02); and for the TW 1100–1300 ms at the left frontal region (p = .02) (see Figure 3). Thus, we found a more negative ERP response to incorrect compared to correct suffixes between 900 and 1300 ms, that is, between 600 and 1000 ms after suffix onset.

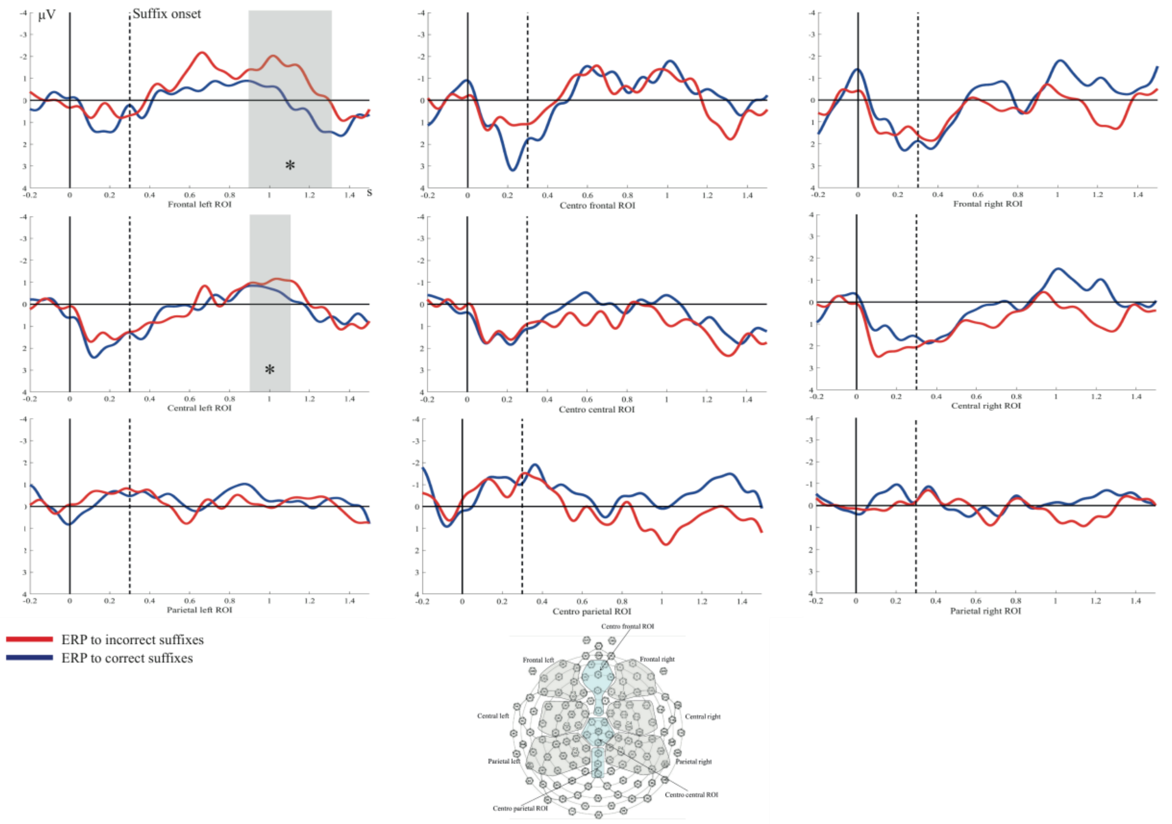

Figure 3. Event-related potentials (ERPs) of the testing phases at day 2. Illustrated are the mean ERPs in response to the correct suffixes containing the nonadjacent dependency (blue line) and to the incorrect suffixes violating the nonadjacent dependency rule (red line) averaged for left, middle, and right frontal, central, and parietal regions of interest (ROIs; see schematic head for details on electrodes). Grey bars and asterisk indicate time windows and regions with significant differences between the ERPs of the two conditions (* p < .05).

3.3 Regression analyses on prediction of behavioral changes by ERPs

We included the following predictors in our regression analysis: 1) the ERP difference wave (incorrect – correct) between 1100 and 1300 ms at the left frontal and at the centro-frontal region at day one (i.e., more positive response to incorrect trials) and 2) the ERP difference wave (incorrect – correct) between 900 and 1300 ms at the left frontal and at the left central region at day two (i.e., more negative response to incorrect trials).

We found a significant linear model fit for the change in behavior from day one to day two [F (2,35) = 5.70; $p < .01$], for which both predictors survived the inclusion criterion of $p = .10$. These were: 1) the ERP effect between 1100 and 1300 ms at day one (i.e., positivity), which was positively associated with the change in task performance from day one to day two and 2) the ERP effect between 900 and 1300 ms at day two (i.e., negativity), which was negatively associated with the change in task performance from day one to day two (see Figure 4). The model could explain 27% of the variance of change in task performance from day one to day two (see Table 1).

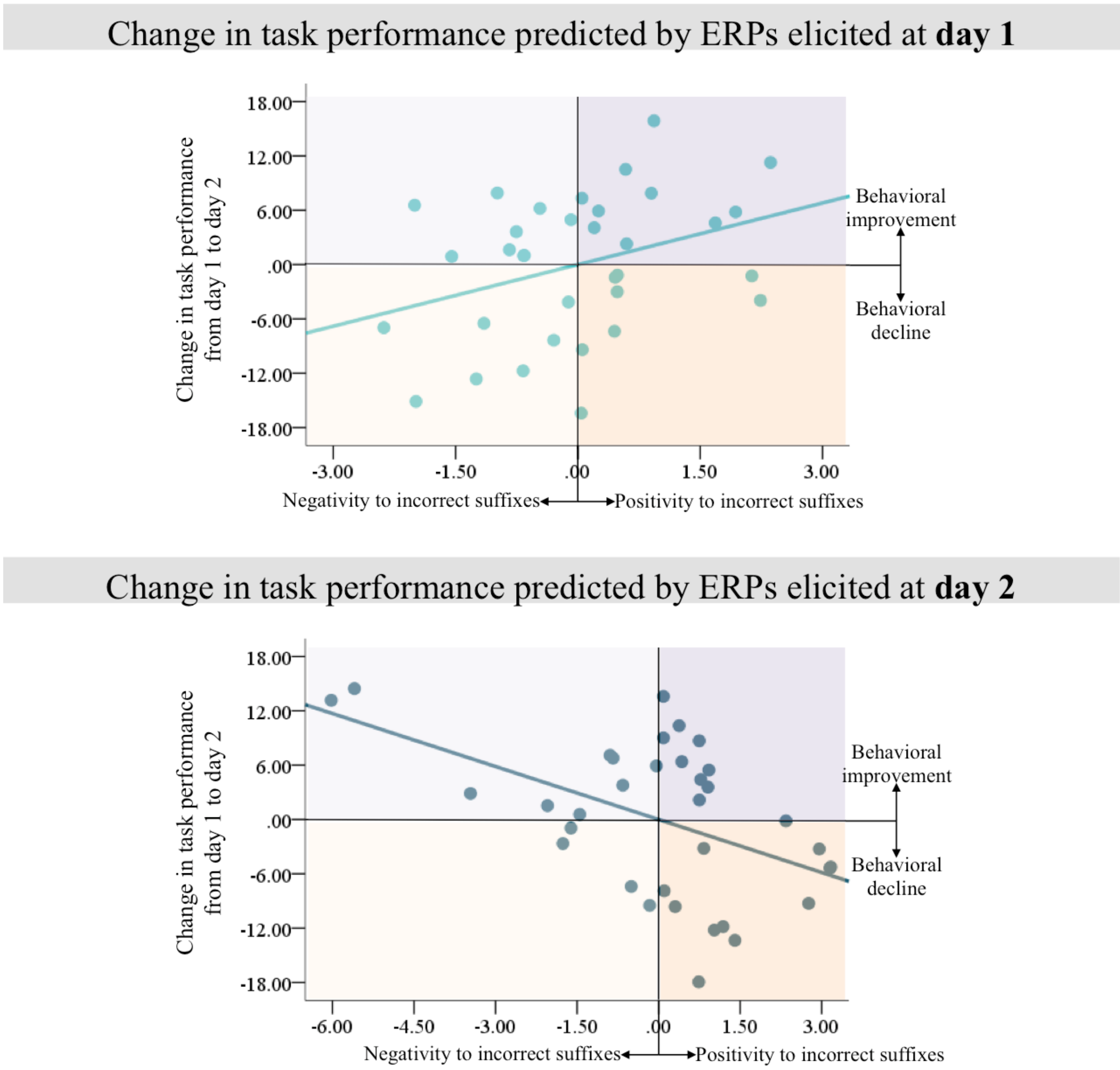

Figure 4. Prediction of change in task performance by event-related potentials (ERPs). The upper part of the figure illustrates how the ERP to incorrect suffixes (i.e., ERP incorrect suffix

– ERP correct suffix) at day 1 is related to the individual change in task performance (i.e., correctly answered trials in percent) from day 1 to day 2 (β = .33; p = .048). The lower part of the figure illustrates how the ERP to incorrect suffixes (i.e., ERP incorrect suffix – ERP correct suffix) at day 2 is related to the individual change in task performance (i.e., correctly answered trials in percent) from day 1 to day 2 (β = –.49; p = .004). Children showing an improvement in behavior are highlighted by blue squares, while children showing a decline in behavior are highlighted by orange squares. Darker shades indicate a more positive ERP to incorrect compared to correct suffixes, while lighter shades indicate a more negative ERP to incorrect compared to correct suffixes. Thus, children showing a more positive ERP to incorrect suffixes at day 1 are more likely to improve behaviorally compared to children showing a less positive ERP at day 1 and children who more strongly improved behaviorally from day 1 to day 2 showed a more negative ERP to incorrect suffixes at day 2 compared to children who less strongly improved behaviorally.

Table 1. *The final model of the linear regression to analyze the association between the significant event-related potential (ERP) effects when contrasting incorrect to correct trials [i.e., more positive ERP to incorrect trials at day one (1100 – 1300 ms) and more negative ERP to incorrect trials at day two (900 – 1300 ms) with change in task performance (i.e., correctly answered trials) from day one to day two.*

				95% Confidence Interval for B	
	B	*SE* B	*β*	Lower bound	Upper bound
Constant	–1.38	1.67		–4.79	2.04
Positivity day one (1100 – 1300 ms)	2.27	1.10	.33*	0.02	4.52
Negativity day two (900 – 1300 ms)	–1.95	0.63	–.49**	–3.24	–0.67

Note. B = unstandardized coefficient; *SE* = Standard error; *β* = standardized coefficient. * $p < .05$; ** $p < .01$

3.4 Post-hoc analyses

Our results indicated a change from a positivity to a negativity in the ERP response to incorrect suffixes from day one to day two, both of which seem to be related to individual changes in task performance from day one to day two. To further support this finding, we calculated the difference between the positivity at day one and the negativity at day two post-hoc, according to the procedure of calculating the change in task performance as described above. A negative value would indicate a stronger change from positivity at day one to negativity at day two. We then calculated Pearson's bivariate correlation coefficient to analyze the association between the changes in ERP polarity from day one to day two and the change in task performance from day one to day two.

The result showed a significant negative correlation between the individual change in ERP polarity from day one to day two and the individual change in task performance from day one to day two ($r = -.52$; $p < .005$).

4. Discussion

The aim of the present ERP study was to investigate NAD-learning by means of recollection using a miniature version of a natural language in 7-year-olds. Specifically, we not only tested NAD processing directly after learning, but also recollection after a retention period involving sleep (i.e., at the next day). On the first day, German-speaking children were exposed to Italian sentences containing NADs (e.g. *La sorella <u>sta</u> cant<u>ando</u>*; the sister <u>is</u> sing<u>ing</u>). Learning phases were followed by testing phases in which participants heard a mixture of correct sentences containing the same NADs as during learning phases, as well as incorrect sentences containing NAD violations (e.g. *La sorella <u>sta</u> cant<u>are</u>*; the sister <u>is</u> sing<u>Ø</u>; see Friederici et al., 2011; Mueller et al., 2009), while they performed a grammaticality judgment task. To then test recollection, participants were re-invited the following day, on which they were presented with testing phases only, while again their EEG data and grammaticality judgments were acquired.

The grammaticality judgment task at either day did not reveal any NAD-learning (above-chance performance) in our group of 7-year-old children, contradicting findings by Raviv and Arnon (2017), who showed behaviorally that 7-year-olds had successfully learned an artificial grammar. This contradiction in results could have one of the following reasons: (1) Our natural language stimuli were more complex than the artificial language involving syllable triplets in Raviv and Arnon (2017)'s study and may thus be more difficult to learn. (2) Our grammaticality judgment task was more difficult than the two-alternative forced choice task used by Raviv and Arnon (2017), in which children were presented with two stimuli, one of which conformed to a familiarized language and one did not. It is conceivable that having the direct comparison between a correct and an incorrect example in the 2-AFC task, including the knowledge that one sentence is correct and one incorrect, facilitates learning compared to the grammaticality judgment task used in the present study. Thus, it might be concluded that the present grammaticality judgment task is still too difficult for 7-year-old children (see also Lammertink Witteloostuijn, Boersma, Wijnen, & Rispens, 2018) such that they cannot successfully show the same behavior as adults (i.e., above chance level, see Mueller et al., 2009). However, when looking at behavioral changes in performance (i.e., correct responses) from day one to day two, we found behavioral changes in the positive direction for some children (i.e., more correct answers at day two compared to day one), possibly indicating NAD-learning at least for some of the 7-year-olds after a retention period involving sleep.

In the following, we will first discuss the ERP findings of NAD processing, before elaborating on the association between behavioral and neurophysiological responses. At the neural level, we found a more positive ERP response to NAD violations (i.e., 800 to 1200 ms after suffix onset) during testing phases on day one. In contrast, we found a more negative

ERP response to NAD violations (i.e., 600 to 1000 ms after suffix onset) during testing phases on day two. Regarding this change in polarity of the ERP effects following overnight-retention, we suggest a change in representation as interpretation. In previous studies using this paradigm, positive ERPs in response to NAD violations have been reported for infants (Friederici et al., 2011) and adults when their PFC was inhibited by transcranial direct current stimulation (Friederici et al., 2013). It has been suggested that infants employ associative learning strategies and that with increased PFC development, learning mechanisms change to controlled top-down learning (Skeide & Friederici, 2016). Because both, infants, whose PFC is not yet fully developed (Huttenlocher, 1990), and adults with a temporarily inhibited PFC, showed a positive ERP response to NAD violations, more positive ERPs have been interpreted to indicate associative learning of NADs (Friederici et al, 2011; 2013). In our study, the positivity on day one may thus indicate that children have formed associative representations of the NADs before retention. In contrast, we found a negative ERP to NAD violations on day two. Negative ERP responses to NAD violations in this paradigm have been reported for adults under standard conditions, that is, when their PFC was not inhibited (Friederici et al., 2013; Mueller et al., 2009; Citron et al., 2011). This negativity was interpreted to reflect an N400, indicating lexical-semantic processing (Mueller et al., 2009; Kutas & Federmeier, 2000). Children’s negativity found in the present study on day two occurred slightly later and more frontally, most likely reflecting an immature N400 (see Hahne, Eckstein, & Friederici, 2004; Henderson, Baseler, Clarke, Watson, & Snowling, 2011), indicating a lexical strategy (Mueller et al., 2009) on testing day two. Taken together, we suggest the positivity in the ERP on day one to indicate associative NAD-learning and the negativity in the ERP on day two to indicate a lexical processing strategy, where both mechanisms might be beneficial for NAD-learning in 7-year-old children.

Indeed, we could show that both, the positivity on day one, as well as the negativity on day two, were associated with children’s behavioral changes in performance from day one to day two. Specifically, children who showed a more positive ERP response to NAD violations at day one and a more negative ERP response to NAD violations at day two showed a stronger behavioral change towards more correct answers at day two compared to day one. These results possibly indicate that 7-year-olds show recollection after a retention period involving sleep and that their representations of NADs change over this retention period. Our post-hoc analysis further supports this assumption by showing that a stronger change from a positive ERP effect at day one to a negative ERP effect at day two was correlated with stronger positive behavioral changes (i.e., more correct answers at day two compared to day one), most likely associated with a change of NAD representation during a retention period enabling 7-year-old's recollection of NADs on the second testing day.

A study by Friedrich and colleagues (2017) offers insight into a possible mechanism underlying this change of representations during a retention period. Specifically, this study demonstrates an effect of sleep on the representation of learned associations between object-word pairs in infants. Object-word pairs were learned through mere phonological associations by infants who had a short nap after familiarization, while infants who had a longer nap built up semantic long-term memory representations of the object-word pairs. Infants who did not sleep between familiarization and test, however, did not show any evidence for learning the object-word pairs. Similarly, it is possible that in our study the retention period between day one and day two allowed those children who had already built up an associative representation of NADs on day one (indexed by a stronger positivity) to consolidate their associations and transfer them to long-term memory. This consolidation may have enabled children to build more robust representations (indexed by a stronger negativity) of the NADs, enabling these children to recollect the NADs on day two, as indicated by more correct answers on day two.

Our results are in line with previous studies reporting a beneficial effect of a consolidation period involving sleep for artificial grammar learning in infants (Gómez et al., 2006; Hupbach et al., 2009) and for word learning in older children (Backhaus et al., 2008). More specifically, the idea that successful performance in an artificial grammar-learning task is only achieved after consolidation involving sleep is in line with a study by Fischer and colleagues (2006). Crucially, only after a consolidation period that involved sleep, but not a consolidation period without sleep, did participants perform significantly above chance in the artificial grammar-learning task (Fischer et al., 2006). Our study provides further support for the beneficial effect of sleep on NAD-learning and recollection.

It has recently been debated whether (artificial) grammar learning is governed by similarity-based or rule-based learning mechanisms (see Hahn & Chater, 1998). Similarity-based learning occurs when (chunks of) familiarized items are memorized and these memorized representations are then compared to test items during testing phases. During rule-based learning, on the other hand, rules underlying the items are formulated and tested on the test items, finally resulting in rule-based representations. Opitz and Hoffmann (2015) provided evidence that both mechanisms play a role in artificial grammar learning, with similarity-based learning being especially prominent in initial stages of learning, while rule-based learning builds up over time. Similarly, in our study, it would be possible that on day one, children used similarity-based learning to associate the elements of the NADs on day one. It is conceivable that rule-based learning took place over the retention period and allowed children to recollect their rule-based representations of NADs on day two. These different processing mechanisms would then account for the change from positivity to negativity in the ERPs. Because the change from positivity to negativity correlates with change in behavior from day one to day two, this might mean that only those children, who learned similarity-based on day one were able to transform their knowledge to rule-based representations on day two. However, in the present study, we can not directly test whether children relied on similarity-based or rule-based representations at a given day and future studies will have to test this claim.

Taken together, the present study demonstrates that 7-year-old children are able to recollect NADs after a retention period including sleep. ERP responses to NAD violations compared to familiarized NADs indicate that representations of NADs changed during a retention period. Before retention, we found a more positive ERP response to NAD violations, which could indicate associative representations of NADs and/or similarity-based learning. In contrast, after a retention period involving sleep, we found a more negative ERP response to NAD violations, indicating that representations had been stored in long-term memory and thus demonstrating recollection of NADs. Importantly, a stronger change from a more positive ERP response on day one to a more negative ERP response on day two was associated with stronger positive changes in behavior (i.e., more correct answers at day two compared to day one). This could indicate that only those children who had built associative representations of the dependencies on day one were able to consolidate these representations and show recollection on day two, as indicated by a positive change in behavior (i.e., more correct answers at day two compared to day one). These results are the first to show children's recollection of NADs embedded in a natural, foreign language after a retention period involving sleep.

Acknowledgements

The Max Planck Society (GS, MP, RM, CM, ADF), the German Research Foundation, project FR 519/20-1 (FOR 2253; GS, MP, CM, ADF), and the Berlin School of Mind and Brain (MP) funded this project. The authors declare to have no conflict of interest. Further, we would like to thank all the participating families for their commitment, as well as Sophia Röthing for her help with setting-up the experiment and Ulrike Barth for her help with data acquisition and her dedicated work with the participating children.

References

Backhaus, J., Hoeckesfeld, R., Born, J., Hohagen, F., & Junghanns, K. (2008). Immediate as well as delayed post learning sleep but not wakefulness enhances declarative memory consolidation in children. *Neurobiology of Learning and Memory*, *89*(1), 76–78.

Citron, F. M. M., Oberecker, R., Friederici, A. D., & Mueller, J. L. (2011). Mass counts: ERP correlates of non-adjacent dependency learning under different exposure conditions. *Neuroscience Letters*, *487*(3), 282–286. https://doi.org/10.1016/j.neulet.2010.10.038

Fischer, S., Drosopoulos, S., Tsen, J., & Born, J. (2006). Implicit learning - explicit knowing: A role for sleep in memory system interaction TT. *Journal of Cognitive Neuroscience*, *18*(3), 311–319. https://doi.org/10.1162/jocn.2006.18.3.311

Friederici, A. D., Mueller, J. L., & Oberecker, R. (2011). Precursors to natural grammar learning: preliminary evidence from 4-month-old infants. *PLoS One*, *6*(3), e17920.

Friederici, A. D., Mueller, J. L., Sehm, B., & Ragert, P. (2013). Language learning without control: the role of the PFC. *Journal of Cognitive Neuroscience*, *25*(5), 814–821.

Friedrich, M., Wilhelm, I., Mölle, M., Born, J., & Friederici, A. D. (2017). The Sleeping Infant Brain Anticipates Development. *Current Biology*, *27*(15), 2374–2380.e3. https://doi.org/10.1016/j.cub.2017.06.070

Frost, R. L., & Monaghan, P. (2016). Simultaneous segmentation and generalisation of non-adjacent dependencies from continuous speech. *Cognition*, *147*, 70-74.

Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science*, *13*(5), 431–436.

Gómez, R. L., Bootzin, R. R., & Nadel, L. (2006). Naps Promote Abstraction in Language-Learning Infants. *Psychological Science*, *17*(8), 670–675.

Gómez, R., & Maye, J. (2005). The developmental trajectory of nonadjacent dependency learning. *Infancy*, *7*(2), 183–206.

Hahn, U., & Chater, N. (1998). Similarity and rules: distinct? exhaustive? empirically distinguishable?. *Cognition, 65(2-3),* 197-230.

Hahne, A., Eckstein, K., & Friederici, A. D. (2004). Brain signatures of syntactic and semantic processes during children's language development. *Journal of Cognitive Neuroscience*, *16*(7), 1302–1318. https://doi.org/10.1162/0898929041920504

Henderson, L. M., Baseler, H. A., Clarke, P. J., Watson, S., & Snowling, M. J. (2011). The N400 effect in children: Relationships with comprehension, vocabulary and decoding. *Brain and Language, 117(2),* 88-99.

Hupbach, A., Gómez, R. L., Bootzin, R. R., & Nadel, L. (2009). Nap-dependent learning in infants. *Developmental Science*, *12*(6), 1007–1012. https://doi.org/10.1111/j.1467-7687.2009.00837.x

Huttenlocher, P. R. (1990). Morphometric study of human cerebral cortex development. *Neuropsychologia*, *28*(6), 517–527.

Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences*, *4*(12), 463–470. https://doi.org/10.1016/S1364-6613(00)01560-6

Lammertink, I., Van Witteloostuijn, M., Boersma, P., Wijnen, F., & Rispens, J. (2019). Auditory statistical learning in children: Novel insights from an online measure. Applied *Psycholinguistics, 40(2),* 279-302.

Luu, P., & Ferree, T. (2005). Determination of the HydroCel Geodesic Sensor Nets' Average Electrode Positions and Their 10 – 10 International Equivalents. *EGI Technical Note*, (October), 1–11.

Makeig, S., Bell, A. J., Jung, T.-P., & Sejnowski, T. J. (1996). Independent Component Analysis of Electroencephalographic Data. *Advances in Neural Information Processing Systems*, (8), 145–151. https://doi.org/10.1109/ICOSP.2002.1180091

Mueller, J. L., Bahlmann, J., & Friederici, A. D. (2008). The role of pause cues in language learning: The emergence of event-related potentials related to sequence processing. *Journal of Cognitive Neuroscience*, *20*(5), 892–905.

Mueller, J. L., Friederici, A. D., & Männel, C. (2012). Auditory perception at the root of language learning. *Proceedings of the National Academy of Sciences*, *109*(39), 15953–15958.

Mueller, J. L., Friederici, A. D., & Männel, C. (2018). Developmental changes in automatic rule-learning mechanisms across early childhood. *Developmental Science*, e12700.

Mueller, J. L., Oberecker, R., & Friederici, A. D. (2009). Syntactic learning by mere exposure -- An ERP study in adult learners. *BMC Neuroscience*, *10*(89).

Nelson, D. G. K., Jusczyk, P. W., Mandel, D. R., Myers, J., Turk, A., & Gerken, L. (1995). The head-turn preference procedure for testing auditory perception. *Infant Behavior and Development*, *18*(1), 111–116.

Newport, E. L., & Aslin, R. N. (2004). Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology*, *48*(2), 127–162.

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, *2011*.

Opitz, B., & Hofmann, J. (2015). Concurrence of rule-and similarity-based mechanisms in artificial grammar learning. *Cognitive psychology, 77*, 77-99.

Peña, M., Bonatti, L. L., Nespor, M., & Mehler, J. (2002). Signal-driven computations in speech processing. *Science*, *298*(5593), 604–607.

Perrin, F., Pernier, J., Bertrand, O., & Echallier, J. F. (1989). Spherical splines for scalp potential and current density mapping, *Clinical Neurophysiology*, 184–187.

Raviv, L., & Arnon, I. (2017). The developmental trajectory of children's auditory and visual statistical learning abilities: modality-based differences in the effect of age. *Developmental Science, 21(4), e12593*

Shufaniya, A., & Arnon, I. (2018). Statistical Learning Is Not Age-Invariant During Childhood: Performance Improves With Age Across Modality. *Cognitive science, 42(8),* 3100-3115.

Skeide, M. A., & Friederici, A. D. (2016). The ontogeny of the cortical language network. *Nature Reviews Neuroscience*, *17*(5), 323.

World Medical Association. (2013). World Medical Association Declaration of Helsinki Ethical Principles for Medical Research Involving Human Subjects. *Jama*, *310*(20), 2191. https://doi.org/10.3917/jib.151.0124