

Modality-specific and modality-independent neural representations work in concert in predictive processes during sequence learning

Teodóra Vékony^{1,†}, Ádám Takács^{2,3,†,*}, Felipe Pedraza^{4,5}, Frederic Haesebaert⁵, Barbara Tillmann^{1,6}, Imola Mihalecz¹, Romane Phelipon¹, Christian Beste^{2,3,†}, Dezso Nemeth^{1,7,8,†}

¹Université Claude Bernard Lyon 1, CNRS, INSERM, Centre de Recherche en Neurosciences de Lyon CRNL U1028 UMR5292, 95 Boulevard Pinel, Bron F-69500, France,

²Cognitive Neurophysiology, Department of Child and Adolescent Psychiatry, Faculty of Medicine, TU Dresden, Fetscherstraße 74, Dresden D-01307, Germany,

³Faculty of Medicine, University Neuropsychology Center, TU Dresden, Fetscherstraße 74, Dresden D-01307, Germany,

⁴Université de Lyon, Université Lyon 2, Laboratory EMC (EA 3082), Bron F-69500, France,

⁵Université Claude Bernard Lyon 1, CNRS, INSERM, Centre de Recherche en Neurosciences de Lyon CRNL U1028 UMR5292, 95 Boulevard Pinel, PSYR2 Team, Bron F-69500, France,

⁶CNRS, UMR5022, Laboratoire d'Etude de l'Apprentissage et du Développement, Université de Bourgogne, Dijon F-21048, France,

⁷Institute of Psychology, ELTE Eötvös Loránd University, Kazinczy u. 23-27, Budapest H-1075, Hungary,

⁸Brain, Memory and Language Research Group, Institute of Cognitive Neuroscience and Psychology, Research Centre for Natural Sciences, Budapest H-1117, Hungary

*Corresponding author: Cognitive Neurophysiology, Department of Child and Adolescent Psychiatry, Faculty of Medicine Carl Gustav Carus, TU Dresden, Fetscherstrasse 74, Dresden 01307, Germany. Email: Adam.Takacs@ukdd.de

†Teodóra Vékony and Ádám Takács contributed equally to this work

‡Senior authors

Probabilistic sequence learning supports the development of skills and enables predictive processing. It remains contentious whether visuomotor sequence learning is driven by the representation of the visual sequence (perceptual coding) or by the representation of the response sequence (motor coding). Neurotypical adults performed a visuomotor sequence learning task. Learning occurred incidentally as it was evidenced by faster responses to high-probability than to low-probability targets. To uncover the neurophysiology of the learning process, we conducted both univariate analyses and multivariate pattern analyses (MVPAs) on the temporally decomposed EEG signal. Univariate analyses showed that sequence learning modulated the amplitudes of the motor code of the decomposed signal but not in the perceptual and perceptual-motor signals. However, MVPA revealed that all 3 codes of the decomposed EEG contribute to the neurophysiological representation of the learnt probabilities. Source localization revealed the involvement of a wider network of frontal and parietal activations that were distinctive across coding levels. These findings suggest that perceptual and motor coding both contribute to the learning of sequential regularities rather than to a neither–nor distinction. Moreover, modality-specific encoding worked in concert with modality-independent representations, which suggests that probabilistic sequence learning is nonunitary and encompasses a set of encoding principles.

Key words: electroencephalography; temporal decomposition; multivariate pattern analysis; sequence learning; statistical learning.

Introduction

Many parts of our lives depend on the predictive processes derived from sequential information, such as playing an instrument or using languages (Conway 2020). However, it remains a key question how the perceptual and motor codes contribute to the detection, encoding, and retrieval of such regularities, and whether neurophysiological coding levels work in an independent or concerted fashion (Frost et al. 2015; Conway 2020).

In the visual modality, sequence items are typically presented continuously, while participants are asked to respond to a stimulus feature (Howard and Howard 1997; Conway 2020). Unknown to them, features change according to a predetermined sequence, and the extraction of the sequential regularities leads to faster responses to predictable than to unpredictable stimuli. The

question is how is learning driven by the representation of the visual sequence (perceptual coding) and by the representation of the response sequence (motor coding) (Deroost and Soetens 2006; Nemeth et al. 2009). Performing a visuomotor sequence task or simply observing the visual stream of regularities can lead to comparable learning effects (Remillard 2003; Song et al. 2008). Therefore, sequence learning is possible without motor coding. However, learning the response sequence can also contribute to the learning performance independently of perceptual processes (Goschke 1998; Conway 2020) as sensory manipulations may affect perceptual but not motor sequence learning (Song et al. 2008). Also, motor learning is more efficient in stabilizing memories than perceptual learning (Hallgató et al. 2013). Visual and motor codes during learning operate

Received: December 28, 2022. Revised: February 20, 2023. Accepted: February 21, 2023

© The Author(s) 2023. Published by Oxford University Press. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

For commercial re-use, please contact journals.permissions@oup.com

as overlapping but separable entities, working in concert with domain-general memory processes that can link stimulus-based perceptual and motor-based response coding systems (Frost et al. 2015; Conway 2020). Learning probabilistic sequences likely relies on the concomitant coding of stimulus, response, and stimulus–response translational information (Takács et al. 2021). The question is not whether perceptual or motor codes contribute to the learning of sequential regularities but how the perceptual and motor codes work in concert.

Interestingly, similar questions were asked recently in the context of stimulus–response binding. As the theory of event coding (TEC) posits (Hommel et al. 2001; Hommel 2019), stimuli and responses are represented by their commonly coded features, such as the color of the stimulus or the identity of the response finger. Binding between stimulus and response features creates event files, a representational format that encompasses both stimulus-level (object files) and response-level (action files) encoding. It was proposed that binding provides building blocks for sequence learning (Eberhardt et al. 2017; Haider et al. 2020; Takacs et al. 2021). Binding between stimulus and response features can occur without a predictive relationship between the 2 (Hommel 2004; Frings et al. 2020). However, if they are embedded in a stream of sequential information, sequence learning can reinforce bindings (Takacs et al. 2021). A series of experiments showed that it was possible to induce interference (Eberhardt et al. 2017) or memory transfer (Haider et al. 2020) between stimulus location (stimulus feature-based) and response location (response feature-based) sequences. Thus, according to behavioral data, sequential regularities were coded not only in a modality-specific fashion but also in abstract feature codes. The functional connections between event file binding and sequence learning (Takacs et al. 2021) pose the question of whether TEC's assumption on the nature of coding levels (Hommel et al. 2001; Hommel 2004; Frings et al. 2020) and the methods to study them (Takacs, Mückschel, et al. 2020; Takács et al. 2022) can be applied to sequence learning.

The simultaneous nature of coding levels presents a methodological challenge when their effects need to be studied separately. This challenge is even more vexing on the level of neurophysiological mechanisms (Ouyang and Zhou 2020). In binding studies (Opitz et al. 2020; Takacs, Mückschel, et al. 2020; Takacs, Zink, et al. 2020; Takacs et al. 2021; Eggert et al. 2022), temporal decomposition was used to disentangle the concomitant coding levels. The separation of simultaneous coding levels is possible in the neurophysiological (EEG) signal by using temporal decomposition, such as the residue iteration decomposition (RIDE) (Ouyang et al. 2015a; Ouyang and Zhou 2020; Takács et al. 2022). RIDE distinguishes between concomitantly coded stimulus-related (S-cluster), response-related (R-cluster), and stimulus–response translational information (C-cluster) in the EEG signal. The decomposed clusters represent coding levels with different latency variability, that is, temporal dynamics that can be locked and closely related to stimulus presentation (S-cluster), response execution (R-cluster), and variable latency that is not locked to markers but detected by template matching (C-cluster) (Ouyang et al. 2011, 2015b; Verleger et al. 2014; Ouyang and Zhou 2020). Previous studies linked the response-related signal to action file processes (Stürmer et al. 2013; Takacs et al. 2021), while the stimulus–response translational information in the EEG reflected event file binding studies (Opitz et al. 2020; Takacs, Mückschel, et al. 2020; Takacs, Zink, et al. 2020; Eggert et al. 2022). Thus, the TEC's predictions could be mapped onto the separation of the neurophysiological coding levels. Importantly, the RIDE data were also successfully used in multivariate pattern analysis (MVPA)

(Takacs, Mückschel, et al. 2020; Eggert et al. 2022) to decode event files in the decomposed clusters. RIDE has been validated against other temporal decomposition methods and has been compared with independent component analysis (ICA) (Ouyang et al. 2015b). Crucially, RIDE showed an advantage when subcomponents partially overlap but originate from different temporal lockings (Ouyang et al. 2015b; Ouyang and Zhou 2020). Therefore, a combination of temporal decomposition and MVPA was chosen as a method to investigate the contributions of perceptual (S-cluster), motor (R-cluster), and abstract or not modality-specific (C-cluster) coding levels in the development of sequential memory representations.

A previous study (Takács et al. 2021) found that event-related potentials (ERPs), both in the stimulus–response translational and stimulus clusters, reflected sequence learning; however, the response cluster did not. Importantly, Takacs et al. (2021) used cues in the sequence to distinguish between intentionally (cued) and incidentally (uncued) learnt information. However, the intention to learn might have altered the weight of perceptual and motor coding in sequence learning (Rüsseler and Rösler 2000). Therefore, probabilistic sequence learning in the current study remained uncued to ensure the learning was incidental (Howard and Howard 1997; Song et al. 2008). Yet, a univariate analysis of the decomposed ERPs presupposes that ERPs are sensitive markers of sequence learning and the effects are focal (i.e. specific to a channel, channel pair, or a pool of channels). As we are not aware of evidence that these requirements are all fulfilled, we also employed a data-driven, multivariate approach. We used a protocol to combine temporal signal decomposition with MVPA to investigate configurations in the neurophysiological signal that are related to the perceptual, motor, and translational codes of sequence learning (Takács et al. 2022).

We expected that sequential regularities could be decoded from the decomposed EEG, and the decoded representation would be stable in all 3 clusters, albeit with distinctive time courses.

Moreover, we compared their neural sources to evaluate the distinctiveness of coding levels (Southwell and Chait 2018; Takács et al. 2021). While sequence learning is rooted in the basal ganglia (Janacek et al. 2019, 2022), it also taps into a wide network consisting of the lateral occipital cortex, angular gyrus, precuneus, anterior cingulate cortex, and superior frontal gyrus (Park et al. 2022). Of note, EEG-based source localization is not suitable to uncover learning-related activity changes in all the structures listed above, particularly in the basal ganglia. We expected learning-related activation modulations in the precuneus, the angular gyrus, the anterior cingulate cortex, the inferior frontal, and the superior frontal gyri (Takács et al. 2021; Park et al. 2022).

Materials and methods

Participants

Participants were right-handed young adults with normal or corrected-to-normal vision. None of the participants reported any neurological or psychiatric disorder at the time of the study. Complete data were available for 44 participants. After preprocessing, 43 participants remained in the final sample (24 female, 19 male, $M_{\text{age}} = 22.46$ years, $SD = 2.89$). All participants provided written informed consent before enrolment and received financial compensation for their participation (20 €/h). The study protocol was approved by the relevant institutional board ("Comité de Protection des Personnes Est I," ID RCB 2019-A02510-57), and the experiment was carried out in agreement with the declaration of Helsinki.

Experimental procedure

The experiment started with 5 min of resting-state EEG recording with eyes open. Resting-state recordings were routinely collected as part of a larger project and have not been analyzed in the current study. During the resting state, a white fixation cross appeared in front of a black background, and participants were instructed to fixate on the cross without moving. Afterward, participants performed the alternating serial reaction time (ASRT) task (Fig. 1). During the task, yellow arrows appeared in the middle of the screen consecutively, and participants were instructed to press a key corresponding to the spatial direction of the upcoming arrow on a Cedrus RB-530 response pad (Cedrus Corporation, San Pedro, CA). Participants were asked to keep their thumbs and index fingers above the 4 response keys. First, participants completed 3 blocks of 85 stimuli each in which arrows pointed in randomly selected directions to familiarize participants with the response keys (practice blocks). After the practice blocks, 25 blocks (85 stimuli/block) of ASRT were completed by the participants (see details below). After a set of 5 blocks, ~5 min of rest was inserted during which EEG impedance was verified and improved if necessary.

Task and stimuli

Probabilistic sequence learning was measured by the ASRT task, modified to fit the EEG measurements. In the task, a yellow arrow appeared at the center of the screen and in front of a black background for 200 ms. The arrow pointed in either the left, up, down, or right direction. After that, a white fixation cross was presented on the screen for 500 ms. If the participant pressed the response key corresponding to the direction of the arrow during this time window, the fixation cross remained on the screen for another 750 ms. To enhance learning, visual feedback was presented for incorrect or missing responses: If an incorrect response or no response was given, an “X” or an exclamation mark was presented on the screen for 500 ms and then the fixation cross appeared for 250 ms. Unbeknownst to the participants, the order of appearance of the arrows follows a predetermined alternating sequence. In this 8-element sequence, pattern (P) and pseudo-random elements (r) alternated (e.g. r-2-r-4-r-3-r-1, where numbers represent a spatial position predetermined by the sequence [1=left, 2=up, 3=down, 4=right], and rs denote a pseudo-random position). With the permutation of the four possible spatial directions, 24 different 8-element alternating sequences can be formed (e.g. r-2-r-1-r-3-r-4; r-1-r-2-r-3-r-4; r-1-r-2-r-4-r-3). However, only 6 unique permutations exist, as, for instance, the sequence of r-2-r-4-r-3-r-1 is indistinguishable from r-3-r-1-r-2-r-4. The triplets that cannot be formed by 2 pattern elements and 1 random element in the middle appear with a lower probability. For instance, 2-1-3, as “cannot be formed” in a P-r-P structure, as the first and the third elements, “2” and “3,” are not 2 consecutive pattern elements in the sequence of r-2-r-4-r-3-r-1. We refer to these types of triplets as “low-probability triplets.” It is important to note that the terms high-probability and low-probability triplets also refer to the predictability of the final (third) element of a triplet: the third element of a high-probability triplet is more predictable from the first element of the triplet than in the case of a low-probability triplet (second-order transitional probabilities).

Each trial of the ASRT task can be categorized as the third element of a high- or low-probability triplet depending on the 2 preceding trials. As a result of the 8-element sequence, 64 different triplets could be formed. If we know the first 2 elements of the triplet (e.g. 2-1), there are 4 possible outcomes for the

third element: The triplet could end in either 1, 2, 3, or 4. Only 1 of the 4 could result in a high-probability triplet: Using the above example (r-2-r-4-r-3-r-1), only the “4” as the third element will result in a high-probability triplet (in that case, the triplet will be 2-1-4). Therefore, from the 64 possible triplets, 16 will be high-probability triplets, and 48 of them will be low-probability triplets. As every other trial results in a high-probability triplet (as every other element is part of the pattern, P-r-P structure), 50% of triplets will be high-probability triplets. However, 1 of the 4 triplets with an r-P-r structure will also be a high-probability triplet ($50\%/4 = 12.5\%$). Therefore, the 16 high-probability triplets will occur 5 times more ($50\% + 12.5\% = 62.5\%$ of the cases) than the 48 low-probability triplets (37.5% of the cases).

Stimuli were presented in blocks of 85 trials. Each block started with 5 warm-up trials (arrows appeared at random positions) and then the 8-element sequence was repeated 10 times. The sequence remained the same for the entire session. Response times and response accuracy were recorded. After each block, participants received feedback on their performance in the given block (mean reaction time [RT] and accuracy). The feedback message remained on the screen for 5,000 ms. Then, a 15,000-ms long mandatory rest was inserted, and participants could continue with the next block whenever they were ready. The task was completed bimanually.

Behavioral data analysis

Due to the more probable appearance of high-probability triplets, participants typically become faster for these types of trials as the task progresses. Thus, we could measure probabilistic sequence learning by the difference between responses to high- versus low-probability trials: A greater difference between high- and low-probability trials indicates greater learning.

Only trials with correct responses were considered for the behavioral analysis. Trills (e.g. 1-2-1) and repetitions (e.g. 1-1-1) were also removed, as participants often show preexisting tendencies for these types of trials (Nemeth, Janacsek, and Fiser 2013; Éltető et al. 2022). The first 7 trials (the first 5 warm-up trials plus the next 2 that are the first 2 elements of the first triplet) were also removed from the analysis. The remaining trials were categorized as the last element of a high- or a low-probability triplet. The trials of the ASRT task were collapsed into 5 larger units (bins) of analysis, each of them consisting of 425 trials (blocks 1–5 vs. blocks 6–10 vs. blocks 11–15 vs. blocks 16–20 vs. blocks 21–25). For each participant and each unit of analysis, we calculated the median RT separately for high- and low-probability trials. The statistical learning score was calculated by subtracting the median RTs for the high-probability trials from the RTs for the low-probability trials.

Task reliability was considered in the analyses. It was reported that the ASRT has larger test-retest reliability than the nonprobabilistic SRT (Stark-Inbar et al. 2017). Moreover, visuomotor ASRT has a larger internal consistency than auditory versions of the task (Arnon 2020). For our analyses, learning scores in the ASRT are considered to be reliable in a neurotypical adult population from the sample size of $n = 21$ (Stark-Inbar et al. 2017). Additionally, learning scores in the ASRT are considered to be reliable from 25 blocks of task length according to internal consistency and split-half reliability measures (Farkas et al. 2023). Both sample size and task length fulfilled the reliability criteria.

EEG recording

EEG was recorded in an electrically shielded and acoustically attenuated room. Sixty-four actiCAP slim active electrodes were

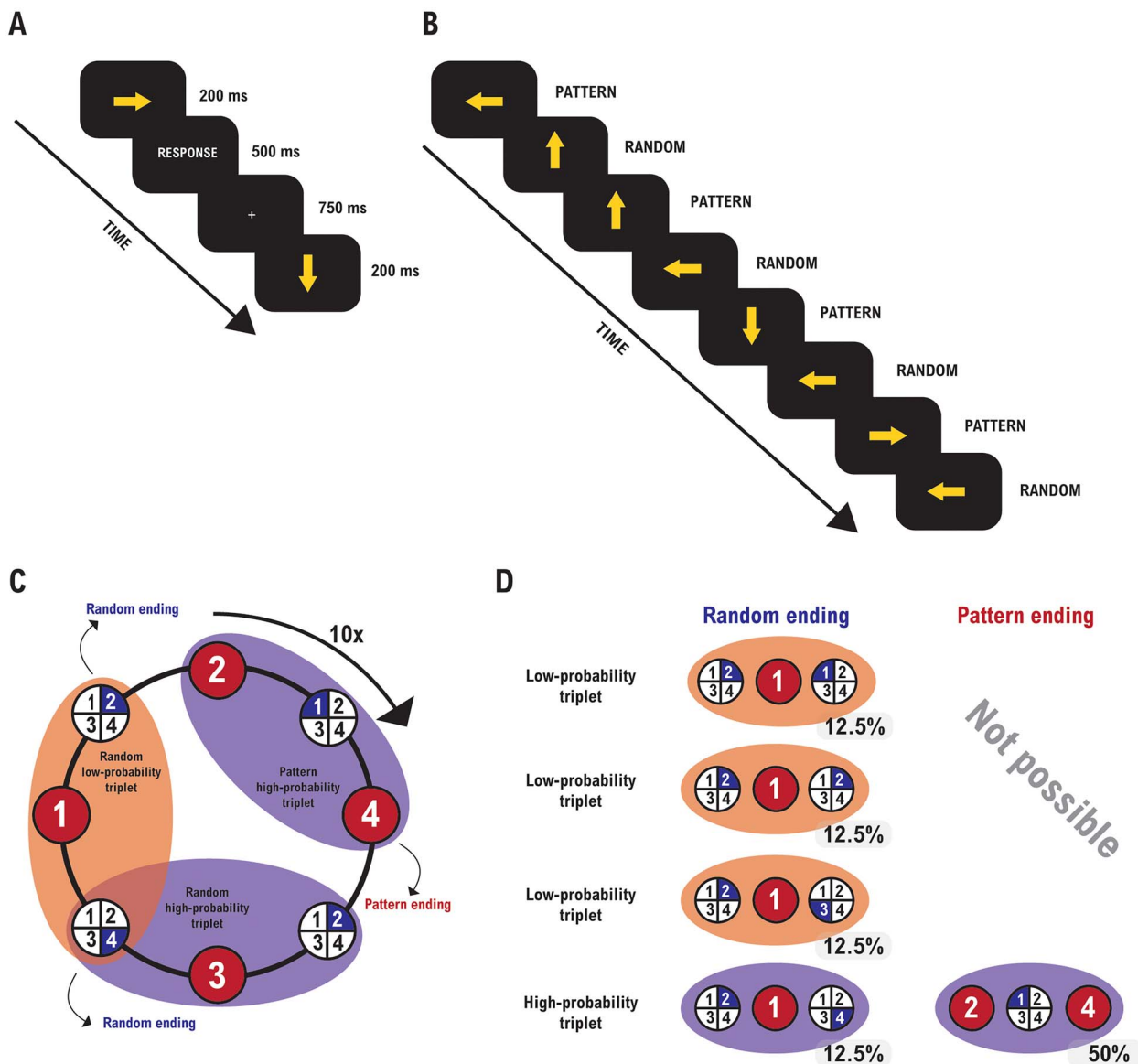


Fig. 1. Task and design. Participants saw an arrow in the middle of the screen, and they were asked to press the response key that corresponds to the direction of the arrowhead. **A)** The timing of the stimulus presentation. **B)** The stimulus presentation followed an 8-element sequence in which pattern and random elements alternated. **C)** As a result of the 8-element sequence, 64 different triplets (runs of 3 consecutive elements) could be formed. Each trial of the ASRT task could be categorized as the third element of a high- or low-probability triplet. Numbers indicate the direction of the arrows (1=left, 2=up, 3=down, 4=right); numbers in large circles represent pattern elements, while numbers in partial circles indicate random elements. **D)** Some triplets were more probable in the task than others. High-probability triplets could either end with a pattern or with random elements, while low-probability triplets always end with a random element.

placed on the scalp according to the international 10–20 system mounted in an elastic cap using BrainAmp DC EEG amplifiers (BrainProducts GmbH, Gilching, Germany). The sampling rate was 500 Hz. The electrode AFz was used as ground, and the reference electrode was placed on the right side of the nose. No online filters were applied. The impedance of the electrodes was kept <25 k Ω .

EEG data preprocessing and segmentation

EEG preprocessing was performed by using Automagic (Pedroni et al. 2019) and EEGLAB (Delorme and Makeig 2004) in Matlab 2019a (The MathWorks Corp.). First, flat channels were removed if necessary, and the data were referenced to an average reference. Next, the PREP preprocessing pipeline (Bigdely-Shamlo et al. 2015) was used to remove line noise at 50 Hz with a multitaper algorithm and remove contamination by bad channels on the average reference. After that, the clean_rawdata pipeline was

used to (i) detrend the data using a FIR high-pass filter of 0.5 Hz (order of 1,286, stop-band attenuation of 80 dB, and transition band of 0.25–0.75 Hz). In this step, flat-line, noisy, and outlier channels were detected and removed. Next, time windows that showed abnormally strong power (>15 SDs relative to calibration data) were reconstructed using Artifact Subspace Reconstruction (burst criterion: 15) (Mullen et al. 2013). Time windows that could not be reconstructed were removed. A low-pass filter of 40 Hz (sinc FIR filter; order: 86) (Widmann et al. 2015) was applied. EOG artifacts were removed using a subtraction method (Parra et al. 2005). Muscular and remaining eye-movement artifacts were automatically classified and removed by using an ICA-based Multiple Artifact Rejection Algorithm (Winkler et al. 2011, 2014). Components reflecting cardiac artifacts were identified using ICLabel (Pion-Tonachini et al. 2019) and were removed consecutively. Finally, all channels that were removed by

Automagic were interpolated using the spherical method. The preprocessed data were segmented according to time-on-task (bin information) and experimental conditions (probability information). Specifically, segments were created separately for high-probability and low-probability conditions throughout the task. Additionally, segments were created separately for each unit of analysis (bin 1: blocks 1–5, bin 2: blocks 6–10, bin 3: blocks 11–15, bin 4: blocks 16–20, and bin 5: blocks 21–25) and high-probability and low-probability conditions within each bin. Segments started –200 ms and ended 750 ms relative to stimulus onset. Only segments with correct responses were included. Current-source density (CSD) transformation with 4 orders of splines was applied to the segmented data (Perrin et al. 1989; Kayser and Tenke 2015). CSD serves as a reference-free spatial filter to highlight scalp topography. Of note, separate datasets were also created without CSD transformation specifically for the source localization analyses. Next, the segmented data were baseline-corrected based on the 200-ms long interval before the stimulus onset.

Residue iteration decomposition

Temporal signal decomposition was performed by using RIDE (Ouyang et al. 2011, 2015a, 2015b; Ouyang and Zhou 2020) in Matlab 2019a (The MathWorks Corp.) as part of the RIDE-MVPA protocol (Takács et al. 2022) based on numerous previous studies (Takacs, Mückschel, et al. 2020; Petruo et al. 2021; Prochnow et al. 2021; Yu et al. 2022). RIDE estimates clusters with different latency information (variable vs. static) and then uses a nested iteration scheme to self-optimize the cluster solution. The procedure is based on segmented single-trial EEG data and was performed on each channel separately. RIDE was used to estimate 3 clusters: C-cluster (“central cluster”) covers intermediate or translational processes between stimulus and response, such as retrieval, decision-making, or response selection; R-cluster (“response cluster”) refers to motor preparation and execution; and S-cluster (“stimulus cluster”) collects information on stimulus-related processes, such as perception and attention (Ouyang and Zhou 2020). In each step of the iteration, the decomposition module estimates S, C, and R (Ouyang et al. 2015b). Specifically, to estimate a cluster, RIDE subtracts the other 2 clusters from the single-trial EEG and aligns the residuals from every trial to the latency of the estimated cluster. As a result, the estimated cluster represents the median waveform. This procedure is iterated until monotonicity is violated. That is, once the estimated latency value changes direction between the iteration steps, convergence is reached, and the final cluster configuration can be analyzed. Several previous studies have shown that RIDE provides a conceptually meaningful separation of concomitant codings in the neurophysiological signal (Wolff et al. 2017; Opitz et al. 2020; Ouyang and Zhou 2020). The approach has been validated in different paradigms, such as oddball (Verleger et al. 2014), flanker (Bluschke et al. 2017), emotional Stroop (Schreiter et al. 2018), event file binding (Friedrich et al. 2020; Opitz et al. 2020; Takacs, Zink, et al. 2020), response inhibition (Stürmer et al. 2013), and response planning tasks (Takacs et al. 2021). The decomposition into 3 clusters requires predefined time windows for the initial cluster estimations. We have selected these time windows based on a previous study that used RIDE in EEG data with a modified ASRT paradigm (Takács et al. 2021): 150–600 ms after stimulus presentation for the C-cluster, the time window between 300 ms before and 300 ms after the response markers for the R-cluster and 0–500 ms after stimulus onset for the S-cluster. Please note, that overlap between the initial search windows for RIDE clusters is a

common practice, as the iterative comparison between cluster solutions was designed with the assumption of overlapping processes (Ouyang et al. 2011, 2015a, 2015b; Ouyang and Zhou 2020). The separation of the clusters is illustrated in the [Supplementary Fig. S1](#). Next, EEG potentials and their topographies were inspected visually in the high- and low-probability conditions separately for the 3 clusters. Based on previous ERP studies with the ASRT paradigm (Kóbor et al. 2018, 2019; Takács et al. 2021), we constrained the search to a frontal/frontocentral negative deviation (N2) and a central/centroparietal positive deviation (P3). Specifically, frontal and frontocentral electrodes were inspected for an N2-like deviation. In the S-cluster, a negative deflection was detected between 240 and 340 ms after stimulus onset with a maximum at electrode FCz (“S-N2”). In the R-cluster, a negative deflection was detected between 320 and 440 ms after stimulus onset with a maximum at electrode FCz (“R-N2”). In the C-cluster, there was no observable N2-like component. Next, central, centroparietal, and parietal channels were inspected for a P3-like deviation. In the C-cluster, a positive deflection was detected between 250 and 380 ms after stimulus onset with a maximum at electrode P3 (“C-P3”). In the R-cluster, a positive deflection was detected between 460 and 580 ms after stimulus onset with a maximum at electrode C2 (“R-P3”). In the S-cluster, there was no observable P3-like component. Within these identified time windows, the mean amplitude was quantified and exported at the single-subject level as an averaged ERP for “univariate” analyses and as single-trial data for “multivariate” analyses.

Univariate analyses

Statistical analyses were performed in JASP 0.16.2. (JASP Team) and followed the procedure of analyzing ASRT data (Howard and Howard 1997; Song et al. 2008). Learning was quantified as a difference between high- and low-probability conditions. That is, shorter RTs for high-probability than for low-probability trials mean learning of the probabilistic sequence information. This learning process was analyzed as the probability (high- vs. low-probability) by time-on-task (blocks 1–5, blocks 6–10, blocks 11–15, blocks 16–20, and blocks 21–25) ANOVAs on the RT and RIDE decomposed mean amplitude data. The Greenhouse–Geisser correction was used when sphericity was violated. Partial eta-squared measure was used as an effect size. Post hoc pairwise comparisons were Bonferroni-corrected.

Multivariate pattern analysis

MVPAs of the RIDE-decomposed data were performed by using the MVPA-light toolbox (Treder 2020) in Matlab 2019a (The MathWorks Corp.). For more details, see the RIDE-MVPA protocol (Takács et al. 2022). First, the classes of high-probability and low-probability across all bins were decoded separately for the 3 RIDE-decomposed clusters to identify the potential neurophysiological representation of stimulus probability. Second, temporal generalization was calculated based on the high- and low-probability class differences in the decomposed clusters to analyze the temporal dynamics and the representational stability of stimulus probability. While in binary decoding, the training and testing were performed on the same time points, in temporal generalization, the time points of the training were also tested on all other time points as well. Since the behavioral analysis showed that the probability effect was not modulated by time-on-task (bins), MVPA was performed on the whole task length. Nevertheless, decoding and temporal generalization analyses separately for each bin and each cluster can be found in [Supplementary Figs. S2–S4](#). The number of trials in the 2 classes

was balanced with undersampling to avoid overfitting (Treder 2020). Classification features consisted of EEG amplitude data at the 64 channels in both stimulus classes. Decoding and temporal generalization were performed for each individual and each RIDE-cluster, while the parameters were kept the same.

For the decoding between high and low probabilities, an L1-support vector machine (SVM) was used as a classifier. SVM performs better than the default linear discriminant analysis if the data are non-Gaussian, noisy, or prone to outliers (Treder 2020). Classifications with SVM were crossvalidated with 5-fold. That is, data were divided into 5 equal parts. In every iteration step, 1 of them was used for testing and the rest for training. After each fold had been used for testing, the average of the iterations was calculated. To identify the time interval with the difference between the high-probability and low-probability classes, the area under the curve (AUC) was used as a measure of decoding performance and was compared to the chance level of $AUC=0.5$. Wilcoxon tests were performed against the chance level for each time point across participants. Cluster-based permutation (1,000 permutations) was used as a method of statistical correction. The time windows indicated by significant decoding at the group level were used to investigate the correlations between neurophysiological classification and behavioral results. AUC values were averaged in significantly above-chance time windows at the single subject level separately for each RIDE cluster. Two-tailed Pearson's correlations between individual AUCs and behavioral measures of sequence learning were reported.

Source localization

Source localization was performed in the standardized low-resolution brain electromagnetic tomography (sLORETA) software package (Pascual-Marqui et al. 2002), which has been shown to provide reliable source estimations coinciding with the TMS and high-resolution MR scanning (Dippel and Beste 2015; Ocklenburg et al. 2018). sLORETA employs a 3-shell spherical head model (MNI152 template) where the intracerebral volume is partitioned into 6,239 voxels using a spatial resolution of 5 mm. The standardized current density is calculated for every voxel of the model. sLORETA provides a single linear solution for the inverse problem without localization bias (Pascual-Marqui et al. 2002; Marco-Pallarés et al. 2005; Sekihara et al. 2005). We used the contrast between high-probability and low-probability conditions for the statistical analyses. Voxel-wise randomization with 5,000 permutations in the statistical nonparametric mapping procedures (SnPM) was used to correct for multiple comparisons. To identify functional neuroanatomical regions showing significant learning effects in the univariate analyses, sLORETA was performed at each time point. In case of a significant probability or bin by probability interaction of the mean amplitude data (see Univariate analyses), the maximum activity within the ERP components time window was selected. To identify the sources of significant learning effects as revealed by the multivariate analyses, time windows with significantly above-chance decoding were averaged for the sLORETA comparable to previous studies (Petruo et al. 2021; Prochnow et al. 2021). This was necessary due to the broad time windows in the MVPA (see Multivariate results).

Results

Behavioral results

Overall accuracy was high in the entire sample (the range was between $93\% \pm 0.006$ in blocks 21–25 low probability condition

to $94\% \pm 0.005$ in blocks 21–25 high-probability condition). Thus, no participant had to be excluded based on low task performance. Learning effects were analyzed on the RT data of correct responses (Fig. 2). The probability (high- vs. low probability) by bin (blocks 1–5, blocks 6–10, blocks 11–15, blocks 16–20, and blocks 21–25) ANOVA on the median RT showed a significant main effect of probability ($F(1, 42) = 35.66, P < 0.001, \eta_p^2 = 0.459$). Participants responded faster to high- ($357.94 \text{ ms} \pm 3.9$) than to low-probability trials ($361.70 \text{ ms} \pm 4.0$). Thus, participants learnt to differentiate between trials based on their nonadjacent transitions. The main effect of bin was also significant ($F(4, 168) = 5.57, \epsilon = 0.677, P = 0.002, \eta_p^2 = 0.117$). Participants were faster in blocks 6–10 ($359.44 \text{ ms} \pm 4.0, P = 0.017$), blocks 11–15 ($358.50 \text{ ms} \pm 3.9, P = 0.002$), blocks 16–20 ($358.23 \text{ ms} \pm 3.7, P = 0.001$), and blocks 21–25 ($358.48 \text{ ms} \pm 4.1, P = 0.002$) than in blocks 1–5 ($364.44 \text{ ms} \pm 4.5$). None of the other pair-wise differences was significant after Bonferroni correction ($P_s > 0.999$). Thus, the general response speed increased between the first phase and the rest of the task irrespective of the probability levels. Finally, the probability by bin interaction was not significant ($F(4, 168) = 1.97, \epsilon = 0.790, P = 0.119, \eta_p^2 = 0.045$).

Neurophysiological results

Univariate results

Learning effects were analyzed on the mean amplitude data separately for the temporally decomposed clusters. The main ERP results are presented in Fig. 3. Additionally, all amplitudes that were used for the statistical tests are displayed in Supplementary Fig. S5.

In the C-cluster, the probability (high vs. low probability) by bin (blocks 1–5, blocks 6–10, blocks 11–15, blocks 16–20, and blocks 21–25) ANOVA on the mean amplitude of the 250–380-ms time window (C-P3) over channel P3 showed that the main effects of probability ($F(1, 42) = 0.07, P = 0.799, \eta_p^2 = 0.002$), bin ($F(4, 168) = 2.66, P = 0.053, \epsilon = 0.721, \eta_p^2 = 0.060$), and the interaction between them were not significant ($F(4, 168) = 0.14, P = 0.945, \epsilon = 0.821, \eta_p^2 = 0.003$).

In the S-cluster, the probability (high- vs. low-probability) by bin (blocks 1–5, blocks 6–10, blocks 11–15, blocks 16–20, and blocks 21–25) ANOVA on the mean amplitude of the 240–340-ms time window (S-N2) over channel FCz showed that the main effect of probability ($F(1, 42) = 3.42, P = 0.071, \eta_p^2 = 0.075$) was not significant. However, the main effect of bin was significant ($F(4, 168) = 3.33, P = 0.012, \eta_p^2 = 0.074$). The mean amplitude of the S-N2 was smaller (less negative) in the blocks 16–20 ($-6.78 \mu\text{V}/\text{m}^2 \pm 1.06, P = 0.029$) and blocks 21–25 ($-6.66 \mu\text{V}/\text{m}^2 \pm 1.06, P = 0.015$) than in blocks 1–5 ($-8.60 \mu\text{V}/\text{m}^2 \pm 1.06$). None of the other pair-wise comparisons between bins was significant ($P_s > 0.304$). The interaction between probability and bin was not significant ($F(4, 168) = 0.91, P = 0.461, \eta_p^2 = 0.021$).

In the R-cluster, the probability (high- vs. low-probability) by bin (blocks 1–5, blocks 6–10, blocks 11–15, blocks 16–20, and blocks 21–25) ANOVA on the mean amplitude of the 320–440-ms time window (R-N2) over channel FCz showed a significant main effect of probability ($F(1, 42) = 5.99, P = 0.019, \eta_p^2 = 0.125$). The R-N2 was larger (more negative) in the high-probability ($-4.03 \mu\text{V}/\text{m}^2 \pm 0.98$) than in the low-probability condition ($-3.13 \mu\text{V}/\text{m}^2 \pm 0.98$). The main effect of bin ($F(4, 168) = 2.59, \epsilon = 0.729, P = 0.058, \eta_p^2 = 0.058$) and the interaction between probability and bin were not significant ($F(4, 168) = 0.41, \epsilon = 0.847, P = 0.769, \eta_p^2 = 0.010$). Thus, more probable trials elicited a larger N2-response than did less probable trials throughout the task. The sLORETA analysis (see Fig. 3)

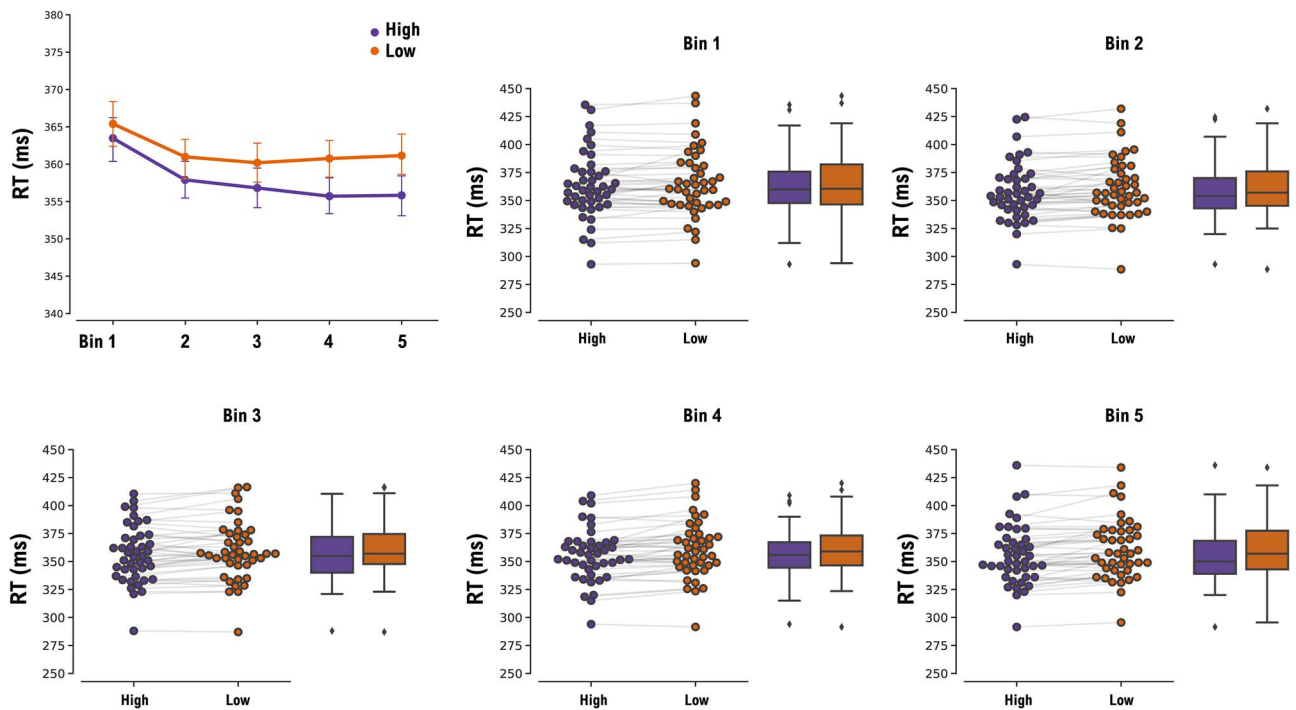


Fig. 2. Behavioral results: individual and group RTs as a function of probability and bin. In the top left corner, group-level RTs are presented as a line graph. Error bars denote the standard error of the mean. Individual RTs are presented as scatterplots and group-level distribution information is visualized as boxplots separately for probability levels in each of the 5 bins (blocks 1–5 vs. blocks 6–10 vs. blocks 11–15 vs. blocks 16–20 vs. blocks 21–25). High: high-probability; low: low-probability.

showed that this probability effect was reflected by activation modulations in the left superior frontal gyrus (BA10).

The probability (high- vs. low-probability) by bin (blocks 1–5, blocks 6–10, blocks 11–15, blocks 16–20, and blocks 21–25) ANOVA on the mean amplitude of the 460–580-ms time window (R-P3) over channel C2 showed a significant main effect of probability ($F(1, 42) = 18.64, P < 0.001, \eta_p^2 = 0.307$). The R-P3 was larger (more positive) in the low probability ($5.43 \mu\text{V}/\text{m}^2 \pm 0.61$) than in the high-probability condition ($4.81 \mu\text{V}/\text{m}^2 \pm 0.61$). The main effect of bin ($F(4, 168) = 0.72, \epsilon = 0.536, P = 0.500, \eta_p^2 = 0.017$) and the interaction between probability and bin were not significant ($F(4, 168) = 1.70, \epsilon = 0.853, P = 0.653, \eta_p^2 = 0.039$). Thus, less probable trials elicited a larger P3 response than more probable trials irrespective of the time-on-task through the bins. The probability effect on the P3 amplitude was related to activation modulations in the middle frontal gyrus (BA6).

Multivariate results

Group-level decoding results

Figure 4A presents the decoding performance between high- and low-probability conditions separately for the C-cluster, S-cluster, and R-cluster data. In the C-cluster, the classification was significantly above chance from between 150 and 680 ms after stimulus presentation. This time window was associated with activation differences in the left precentral gyrus (BA6). In the R-cluster, the classification was significantly above chance between 240 and 620 ms after the stimulus onset. This time window was associated with activation in the left precuneus (BA19) and the left superior parietal lobule (BA7). In the S-cluster, high- and low-probability conditions were decoded above-chance level between 100 and 640 ms relative to stimulus presentation. This period was associated with the right superior frontal gyrus (BA9). Thus, in all 3 RIDE-clusters, decoding analyses provided

successful classifications, suggesting that neurophysiological representations of stimulus probability can be observed at all 3 decomposed coding levels. Notably, classifications in the 3 decomposed signal clusters occurred in partially overlapping time windows. Therefore, source localization analyses were also performed in parts of the time windows that were either specific to a single cluster or shared with at least another one (see Fig. 4B). Specifically, the period of 100–150 ms was decoded above-chance level only in the S-cluster. This time window was associated with the right medial frontal gyrus (BA9). The next period, between 150 and 240 ms, was decoded both in the S-cluster and the C-cluster. In the S-cluster, it was associated with the left precentral gyrus (BA6), while in the C-cluster, it was associated with the right inferior frontal gyrus (BA9). The remaining significant above-chance time windows of the S-cluster (240–640 ms) and the C-cluster (240–680 ms) showed activation differences in the left postcentral gyrus (BA2) and the left precentral gyrus (BA6), respectively.

Temporal generalization results

Figure 4C presents the temporal generalization results separately for the C-cluster, S-cluster, and R-cluster data. In the C-cluster, decoding accuracy was highest (i.e. $\text{AUC} > 0.7$) along the diagonal between 200 and 600 ms after stimulus presentation. Moving from the axis, the above-chance classification gradually decreased but remained significant except for the edges of the matrix. Importantly, the training window of 300–400 ms (i.e. the period corresponding to response selection and execution response, see Behavioral results) generalized to the entire trial length of the test time points. In the S-cluster, decoding accuracy was highest (i.e. $\text{AUC} > 0.7$) along the diagonal between 200 and 480 ms after stimulus presentation. This diagonal classification was not only shorter than in the C-cluster, but the related off-diagonal

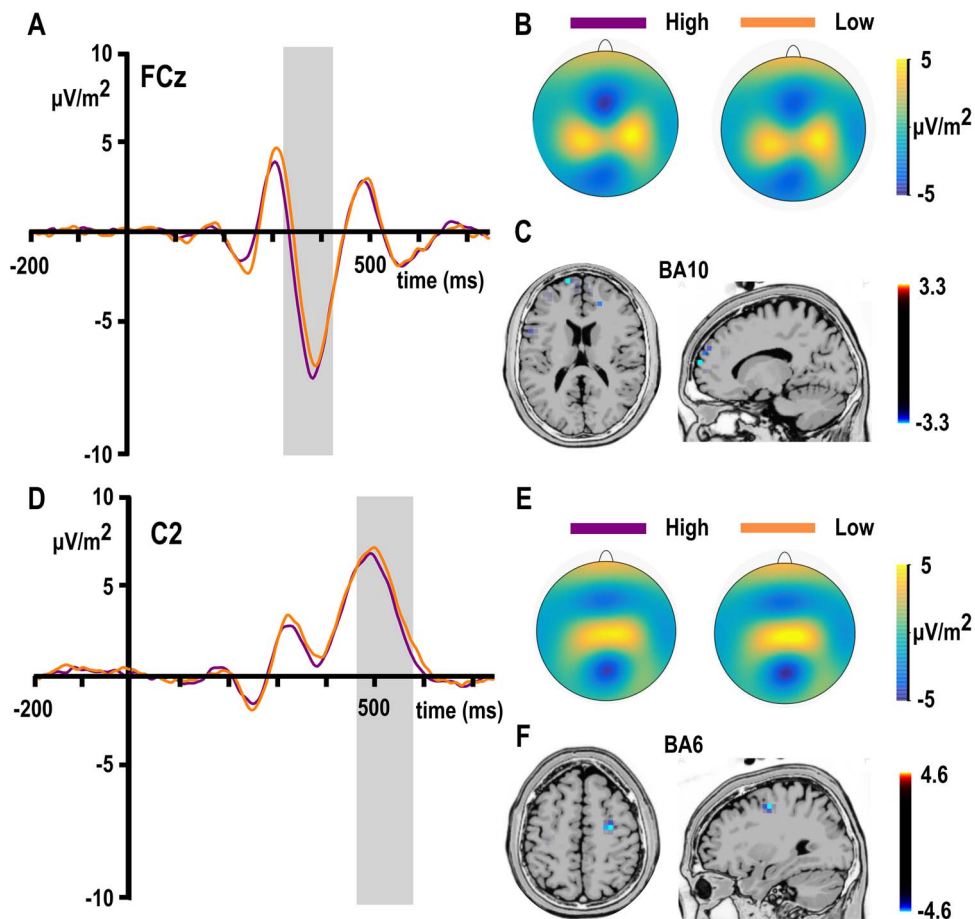


Fig. 3. Univariate results: ERPs and source localization. A) N200 waveforms on channel FCz. Time point 0 represents the stimulus presentation. The analyzed time window (320–420 ms) is marked with a shaded area. The N200 is organized into 2 conditions: high-probability and low-probability. B) The scalp topography plots show the distribution of the mean activity of the two probability levels. C) Results of the sLORETA source localization based on the analyzed time window. The scales denote t-values. D) P3 waveforms on channel C2. Time point 0 represents the stimulus presentation. The analyzed time window (460–580 ms) is marked with a shaded area. The P3 is presented as a function of probability levels: high-probability and low-probability. E) The scalp topography plots show the distribution of the mean activity of the high- and low-probability conditions. F) Results of the sLORETA source localization based on the analyzed time window. The scales denote t-values.

classification was also less pronounced, suggesting a more transient probability representation at the stimulus coding level than at the stimulus–response translational level. Additionally, 2 off-diagonal decoding patterns were observed in the S-cluster that were not the direct continuation of the diagonal one. Namely, the training period of ~100–300 ms was generalized to the test period of ~550–700 ms. Similarly, the training period of ~550–700 ms was generalized to the test period of ~100–300 ms. Finally, in the R-cluster, high decoding accuracy ($AUC > 0.7$) was observed along the diagonal between ~250 and 600 ms that gradually decreased toward the edges of the matrix. Importantly, the training time window of 300–500 ms, which corresponds to the average response time in the task, generalized between 100 and 700 ms in the test time points. Thus, decoding between probability levels showed distinctive generalizability in the 3 decomposed clusters. In all 3 analyses, decoding performance was the highest along the diagonal when training and testing were performed on the same time points. This central pattern was shortest in the S-cluster and longest in the C-cluster. Similarly, off-diagonal generalization was shortest in the S-cluster and longest in the C-cluster.

Connecting the levels of neurophysiological decoding and behavior

Since training time windows in periods that are implicated in response preparation and execution showed long-lasting

generalizations both in the C- and R-cluster data, we have analyzed the possible linear relationship between classification accuracy at the individual level and behavioral learning scores. In the C-cluster, summed learning across bins ($r = -0.297$, $P = 0.056$) and change of learning between the first and the last bins (blocks 1–5 vs. blocks 21–25) ($r = 0.024$, $P = 0.881$) did not correlate significantly with decoding accuracy. Similarly, in the S-cluster, summed learning ($r = -0.203$, $P = 0.196$) and change of learning ($r = -0.175$, $P = 0.268$) did not correlate significantly with decoding accuracy. In the R-cluster, decoding accuracy did not correlate significantly with summed learning ($r = -0.276$, $P = 0.077$), however, the correlation was significant with the change of learning ($r = 0.313$, $P = 0.044$). Thus, participants who showed a bigger change in learning between the beginning and the end of the task (i.e. a steeper learning curve) were better classified based on the R-cluster data.

Discussion

This study investigated the concomitant coding of sequential regularities in perceptual, motor, and perceptual-motor information of the neurophysiological signal. Previous studies suggested that modality-specific encoding in either the perceptual or motor domains is sufficient to develop the representation of sequential regularities (Song et al. 2008), while other accounts assumed

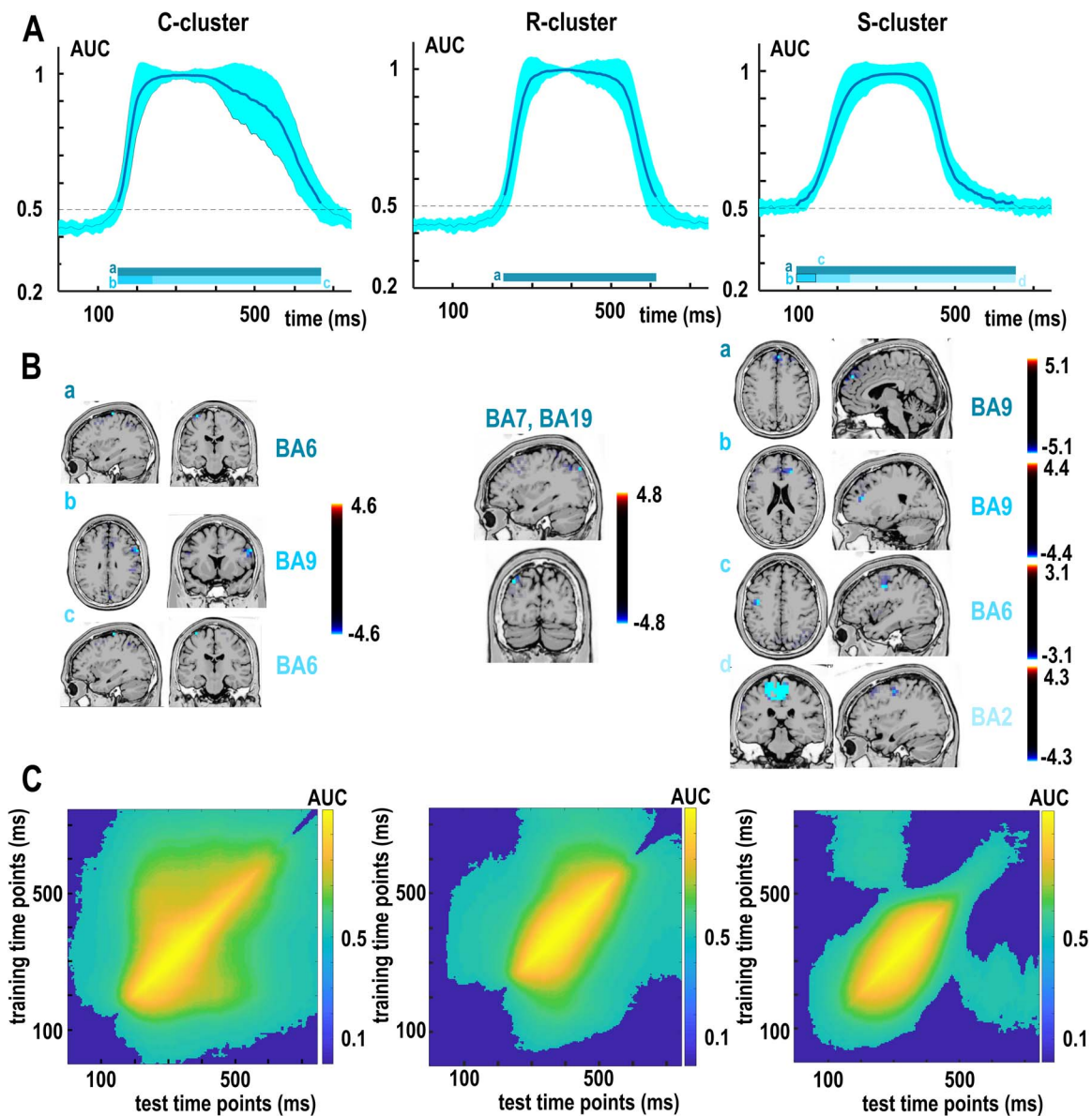


Fig. 4. Multivariate results: decoding, temporal generalization, and source localization. A) Classification performance between high- and low-probability classes represented as AUC separately for the RIDE decomposed C-, R-, and S-cluster EEG data. Time 0 denotes the presentation of the target stimulus. Thicker lines indicate significant time windows ($P < 0.05$; 2-sided cluster-based permutation). The earliest decoding occurred in the S-cluster (100 ms–640 ms). The C-cluster decoding started later and lasted longer (150–680 ms). Finally, in the R-cluster, decoding was significantly above chance for the shortest period (240–620 ms) among the 3 clusters. B) Results of sLORETA source localization for significant time windows as indicated by letters (a–c) in the AUC plots. Color scales denote t -values. C) Temporal generalization matrices separately for the RIDE decomposed C-, R-, and S-cluster EEG data. The plots show the degree to which the classifier when trained on a given time point (y-axis) generalizes to time points in the trial (x-axis). The scales indicate the classifier performance. The diagonal (bottom left to top right) shows classification performance when the classifier is trained and tested on the same time points. C-cluster (left): Sustained decoding pattern along the diagonal extended to off-diagonal parts of the matrix in a jittered fashion. R-cluster (middle): Similar sustained pattern but with shorter off-diagonal extensions. S-cluster (right): A single sustained activation around the diagonal that extended to the off-diagonal parts of the matrix with jittered edges. Additionally, 2 off-diagonal decoding patterns were also observed that suggest generalization between early (100–300 ms) and late parts (550–700 ms) of the trial’s time window.

the existence of a simultaneous modality-independent coding principle (Frost et al. 2015; Bogaerts et al. 2022). Participants performed a visuomotor sequence learning task in which nonadjacent relations predicted the upcoming target either with high or low probability (Howard and Howard 1997). A combination of temporal decomposition and multivariate decoding methods (Takács et al. 2022) revealed the representation dynamics of probability information at the perceptual (stimulus-related), motor (response-related), and central (stimulus–response translational) levels. We found evidence of learning in the behavioral data and

the decomposed motor cluster ERPs but not in the perceptual or central levels. MVPA confirmed that probabilistic regularities can be decoded based on perceptual and translational codes too. Thus, motor coding does not have a privileged role in the neurophysiology of sequence learning. Instead, both perceptual and response codes contribute to the neurophysiological coding of sequential regularities. Moreover, the C-cluster decoding results confirmed an abstract, modality-independent coding of sequence information. The perceptual and translational contributions to probability representations were revealed by using the

multivariate protocol (Takács et al. 2022) but would have remained hidden with univariate tools.

Motor coding alone seemed to be sufficient for incidental sequence learning since, in the univariate analyses, the probability effect was observed only in the R-cluster. Interestingly, in a previous study (Takács et al. 2021), when visuomotor sequences were learnt either incidentally or intentionally, N2 and P3 amplitudes in the S-cluster and the C-cluster showed learning effects but not in the R-cluster. It was suggested that the incidental/implicit nature of the task could switch learning toward motor encoding as opposed to perceptual-motor encoding in intentional/explicit learning (Rüsseler and Rösler 2000). Thus, it is possible that, due to the incidental learning scenario in the current study, only the R-cluster was sensitive enough to detect learning-related ERP differences. Specifically, the R-N2 was larger in high-probability than in low-probability conditions, which likely reflects the more pronounced motor activation before high-probability than before low-probability stimuli (Mückschel et al. 2017). Additionally, the R-P3 was larger in the low-probability than in the high-probability stimuli. This corroborates a previous ERP study (Kóbor et al. 2019) that used the same paradigm: They found larger amplitudes for less probable than for more probable trials both in the undecomposed stimulus-locked and response-locked P3s. It is conceivable that the R-P3 in the current study implies the same probability effect as the response-locked undecomposed P3 (Kóbor et al. 2019). Namely, larger P3 amplitude in the low-probability condition reflects the increased effort needed to retrieve the less frequently used response association. To sum up, the learning effect was observed in the motor code ERPs but not when stimulus information was involved (C- and S-clusters), suggesting that the motor modality alone is sufficient to develop probability representations in visuomotor sequence learning.

However, the multivariate analyses drew a different picture. Classification based on the high- and low-probability trials was successful in all 3 clusters: Significant above-chance decoding was observed in partially overlapping time windows between C-, R-, and S-cluster analyses. Thus, probability levels were represented not only in the motor code but also in the perceptual and central levels. Moreover, the earliest decoding was detected in the S-cluster, which suggests that probability representation is accessible as early as 100 ms after the stimulus onset. Additionally, the perceptual coding preceded the motor 1 with 120 ms. Importantly, before the motor decoding, but after the perceptual time window, the C-cluster showed a sustained representational pattern (starting from 150 ms). As the C-cluster is not related to modality-specific coding, this raises the possibility of a modality-independent coding of probability information with an early onset of accessibility. Curiously, overlapping representations between clusters resemble both decoding evidence (Takács, Mückschel, et al. 2020) and theoretical accounts (Hommel 2004; Takács et al. 2021) of event file coding. In event file coding, C-cluster decoding was suggested to reflect the abstract feature codes, which enables generalization to overlapping events and modality-specific representations of stimulus and response features (Takács, Mückschel, et al. 2020; Eggert et al. 2022). Testing TEC's predictions in sequence learning yielded new insights both in the past (Eberhardt et al. 2017; Haider et al. 2020) and in the current study. Similarly, distinguishing between the roles of binding per se and retrieval of binding-related memory traces became a central topic in event file coding (Frings et al. 2020; Hommel and Frings 2020; Eggert et al. 2022). Connecting the fields of event file coding and sequence learning has potential for both research communities (Takács et al. 2021). Considering the

functional relevance of event files in sequence learning, C-cluster decoding in the current study also suggests an abstract, modality-independent coding level.

Decoding in all 3 clusters is in line with the notion that learning probabilistic information is not a unitary phenomenon but a set of concerted mechanisms: Some of them are subject to modality-specific constraints, while some operate in a domain-general manner (Frost et al. 2015; Bogaerts et al. 2022). Moreover, these coding levels work simultaneously and in overlapping time windows. This nested structure suggests information sharing between coding levels (Bogaerts et al. 2022). As the overlap between the domain-general (C-cluster) and perceptual (S-cluster) coding preceded the motor time window (from 240 ms), abstract probability representations could develop without the contribution of motor coding (Song et al. 2008). Nevertheless, in the current visuomotor experiment, the observed generalization patterns were similar across the 3 clusters: A sustained decoding pattern along the diagonal extended to off-diagonal parts of the matrix in a jittered fashion. Thus, the onset and offset of the representational patterns were gradual. In sum, sequence learning requires a concerted effort between parallel coding of modality-dependent and modality-independent codes (Frost et al. 2015; Conway 2020; Bogaerts et al. 2022).

The current study is in line with the multifaceted nature of learning (sequentially coded) statistical information (Thiessen and Erickson 2013; Daltrozzo and Conway 2014; Arciuli 2017; Arciuli and Conway 2018; BCs et al. 2021; Maheu et al. 2022). It was proposed that the ability to learn probabilistic information relies on partially distinguishable multiple components, and each of them can tap differently into encoding, consolidation, and abstraction of memories (Arciuli 2017; Arciuli and Conway 2018). Understanding the mechanistic role of these components is particularly challenging, given their functional and temporal overlap. The simultaneous modality-specific and modality-independent encoding could have contributed to the difference between univariate and multivariate results. Similarly, the univariate-multivariate mismatch might reflect the difference between focal (electrode-specific ERPs) and spatially not restricted (all electrodes were used as features in the MVPA) effects. Furthermore, the lack of ERP effects in the S- and C-clusters highlights the limitations of univariate methods to uncover EEG signatures of probabilistic sequence learning. We propose that the combined RIDE-MVPA protocol could be used in future studies to uncover remaining questions on the coding of probability information. For instance, how do the S-cluster and C-cluster dynamics evolve if the sequence is observed but no motor response is required? How do the decomposed coding levels contribute to the consolidation, reactivation, or rewiring of the memory traces?

The current study investigated the roles of visual and motor modalities in probabilistic sequences. The presented method can be applied to other crossmodal designs as well. The possibility of transfer between auditory and visual/visuomotor sequence representations is still contentious (Kemény and Lukács 2019; Conway 2020; Han and Reber 2021; Feng et al. 2023). Furthermore, analyzing neurophysiological coding levels of incidentally acquired memories from different paradigms would dovetail our understanding of how encoding and retrieval of procedural representations might differ across learning scenarios. Different tasks and even different measures from the same task can lead to diverging results (Lukács and Kemény 2015; Arciuli 2017; Arciuli and Conway 2018; Takács et al. 2018; BCs et al. 2021; Takács et al. 2021). The ASRT was selected in the current study as it (i) has

good reliability indices (Stark-Inbar et al. 2017; Arnon 2020; Farkas et al. 2023); (ii) provides information on the temporal dynamics of learning (Howard and Howard 1997; Nemeth, Janacsek, and Fiser 2013; Éltető et al. 2022); and (iii) is well adapted to an EEG setting (Kóbor et al. 2018, 2019; Takács et al. 2021).

Interestingly, the temporal aspect of learning (i.e. the learning curve) has contributed to the results only partially. According to the RT data, learning occurred early in the task and did not show significant development later. This was also reflected by the time-invariant results of the univariate ERP and the multivariate decoding analyses (see [Supplementary Figs. S2–S4](#)). However, the significant correlation between MVPA and behavioral data revealed that participants with a steeper learning curve were better classified based on the R-cluster data. Thus, individual differences in the behavior could be mapped to classification of the motor code but not to the perceptual or central coding. This pattern could be explained by the similarity between RT-based learning scores and RT-informed decomposition of the R-cluster (Petruo et al. 2021). Alternatively, the perceptual and central decodings might be related to different facets, such as consolidation or abstraction. It was recently suggested that cross-modal transfer between sequences require an abstract temporal template (Feng et al. 2023), which might be reflected by the central coding level. Conversely, given the C-cluster's prominent role in inhibition (Bluschke et al. 2017; Mückschel et al. 2017; Schreiter et al. 2018; Opitz et al. 2020; Prochnow et al. 2021), it is feasible that behavioral correlations would show in the presence of competing sequences. Further studies needed to understand how different coding levels contribute to individual differences not only in learning but also in consolidation, memory transfer, and memory interference.

The importance of distributed processing was further corroborated by the source localization results, which implicated a wider network of frontal and parietal activations. Frontal sources were identified in the early time windows (medial frontal gyrus and precentral gyrus) and the complete span (superior frontal gyrus) of S-cluster decoding and in the early (inferior frontal gyrus) and entire periods (precentral gyrus) of the C-cluster decoding. The widespread frontal activity in stimulus (S-cluster) and stimulus-response (C-cluster) coding levels could be related to the inverse relationship between frontally rooted executive functions and probabilistic sequence learning (Nemeth, Janacsek, Polner, et al. 2013; Janacsek et al. 2015; Ambrus et al. 2020; Horváth et al. 2022; Park et al. 2022). The superior frontal gyrus was suggested as one of the main hubs of sequence learning (Park et al. 2022). It was proposed that decoupling between superior frontal and parietal regions is crucial for learning (Tóth et al. 2017; Lum et al. 2022; Park et al. 2022). Moreover, the inferior frontal gyrus has been previously linked to the detection and learning of nonadjacent probabilistic regularities (Barascud et al. 2016; Southwell and Chait 2018; Takács et al. 2021). Activation in the inferior frontal gyrus was localized on the right side. This supports the notion of right hemisphere advantage in the processing of visuomotor sequences (Janacsek et al. 2015; Takács et al. 2021). Curiously, parietal sources were linked only to the R-cluster. It was suggested (Gottlieb 2007; Gottlieb and Snyder 2010) that the parietal cortex is essential in binding stimulus and response information and plays an important part in rule-dependent response selection processes. This parietal role is highlighted when the bottleneck of response selection faces a challenging environment (Stock et al. 2014; Vahid et al. 2022). From the 2 parietal sources, the precuneus has also been identified as a hub in the sequence learning-network (Park et al. 2022). This result in the R-cluster data might

be surprising when considering that BA19 is mainly related to visual processes (Kanwisher and Yovel 2006). However, please note that the precuneus was implicated as part of a larger parietal signal that also encompassed the BA7. As opposed to the parietal sources in the R-cluster decoding, the R-N2 and R-P3 deflections were related to frontal activation differences (superior frontal and middle frontal gyrus, respectively). The difference between neural sources of univariate and multivariate effects within the same cluster corroborates the idea that the 2 approaches tapped into different neurophysiological signatures of sequence learning (Southwell and Chait 2018; Takács et al. 2021).

Conclusion

In sum, we have shown evidence that perceptual, motor, and perceptual-motor coding levels all contribute to the neurophysiology of probabilistic sequence learning. The role of the 3 coding mechanisms could be uncovered with multivariate methods only due to the overlap between the motor (R-cluster), perceptual (S-cluster), and translational (C-cluster) representations. In all 3 cases, the neurophysiological representations of probability levels were stable, sustained, and could be generalized to other time points as well. Thus, neither perceptual nor motor encoding have privileged roles in sequence learning (Song et al. 2008; Conway 2020; Takács et al. 2021). Instead, probability can be represented simultaneously in multiple coding levels, including a not-modality-specific one (C-cluster). Successful decoding at the perceptual-motor level suggests that modality-specific encodings work in concert with more abstract, modality-independent sequence learning (Conway 2020).

Data sharing and code accessibility

The code for generating the learning scores from the raw behavioral data, the reported behavioral data, ERP mean amplitude data, individual and group-level MVPA performance data, and codes for the analyses are available via the Open Science Framework (<https://osf.io/5s34c/>). EEG data in various formats (raw, preprocessed, segmented, etc.) will be provided upon request. The study has not been preregistered.

Acknowledgments

The authors are grateful Jeffrey Maillard for managing participant enrolment.

Authors' contributions

Teodóra Vékony (Conceptualization, Data curation, Formal analysis, Investigation, Project administration, Validation, Visualization, Writing—original draft, Writing—review & editing), Ádám Takács (Formal analysis, Funding acquisition, Methodology, Visualization, Writing—original draft, Writing—review & editing), Felipe Pedraza (Data curation, Investigation, Writing—review & editing), Frederic Haesebaert (Methodology, Writing—review & editing), Barbara Tillmann (Conceptualization, Writing—review & editing), Imola Mihalecz (Data curation, Project administration), Romane Phelipon (Data curation, Project administration), Christian Beste (Methodology, Supervision, Writing—review & editing), and Dezsó Nemeth (Conceptualization, Funding acquisition, Resources, Supervision, Writing—original draft, Writing—review & editing)

Supplementary material

Supplementary material is available at *Cerebral Cortex* online.

Funding

This research was supported by the IDEXLYON Fellowship of the University of Lyon as part of the Programme Investissements d'Avenir (ANR-16-IDEX-0005 to DN); National Brain Research Program (project 2017-1.2.1-NKP-2017-00002). Project no. 128016 has been implemented with the support provided by the Ministry of Innovation and Technology of Hungary from the National Research, Development, and Innovation Fund, financed under the NKFI/OTKA K funding scheme (to DN) and the Deutsche Forschungsgemeinschaft (DFG) TA1616/2-1 (to AT).

Conflict of interest statement: None declared.

References

- Ambrus GG, Vékony T, Janacsek K, Trimborn ABC, Kovács G, Nemeth D. When less is more: enhanced statistical learning of non-adjacent dependencies after disruption of bilateral DLPFC. *J Mem Lang*. 2020;114:104144.
- Arciuli J. The multi-component nature of statistical learning. *Philos Trans R Soc B Biol Sci*. 2017;372:20160058.
- Arciuli J, Conway CM. The promise—and challenge—of statistical learning for elucidating atypical language development. *Curr Dir Psychol Sci*. 2018;27:492–500.
- Arnon I. Do current statistical learning tasks capture stable individual differences in children? An investigation of task reliability across modality. *Behav Res Methods*. 2020;52:68–81.
- Barascud N, Pearce MT, Griffiths TD, Friston KJ, Chait M. Brain responses in humans reveal ideal observer-like sensitivity to complex acoustic patterns. *Proc Natl Acad Sci*. 2016;113:E616–E625.
- Bigdely-Shamlo N, Mullen T, Kothe C, Su K-M, Robbins KA. The PREP pipeline: standardized preprocessing for large-scale EEG analysis. *Front Neuroinformatics*. 2015;9:16.
- Bluschke A, Chmielewski WX, Mückschel M, Roessner V, Beste C. Neuronal intra-individual variability masks response selection differences between ADHD subtypes—a need to change perspectives. *Front Hum Neurosci*. 2017;11:
- Bogaerts L, Siegelman N, Christiansen MH, Frost R. Is there such a thing as a 'good statistical learner'? *Trends Cogn Sci*. 2022;26:25–37.
- Conway CM. How does the brain learn environmental structure? Ten core principles for understanding the neurocognitive mechanisms of statistical learning. *Neurosci Biobehav Rev*. 2020;112:279–299.
- Daltrozzo J, Conway CM. Neurocognitive mechanisms of statistical-sequential learning: what do event-related potentials tell us? *Front Hum Neurosci*. 2014;8:437.
- Deroost N, Soetens E. Perceptual or motor learning in SRT tasks with complex sequence structures. *Psychol Res*. 2006;70:88–102.
- Delorme A, Makeig S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J Neurosci Methods*. 2004;134:9–21.
- Dippel G, Beste C. A causal role of the right inferior frontal cortex in implementing strategies for multi-component behaviour. *Nat Commun*. 2015;6:6587.
- Eberhardt K, Esser S, Haider H. Abstract feature codes: the building blocks of the implicit learning system. *J Exp Psychol Hum Percept Perform*. 2017;43:1275–1290.
- Eggert E, Takacs A, Münchau A, Beste C. On the role of memory representations in action control: neurophysiological decoding reveals the reactivation of integrated stimulus-response feature representations. *J Cogn Neurosci*. 2022;34(7):1246–1258.
- Éltető N, Nemeth D, Janacsek K, Dayan P. Tracking human skill learning with a hierarchical Bayesian sequence model. *PLoS Comput Biol*. 2022;18:e1009866.
- Farkas BC, Tóth-Fáber E, Janacsek K, Nemeth D. A process-oriented view of procedural memory can help better understand Tourette's syndrome. *Front Hum Neurosci*. 2021;15:
- Farkas BC, Krajcsi A, Janacsek K, Nemeth D. The complexity of measuring reliability in learning tasks: an illustration using the alternating serial reaction time task. *Behav Res Methods*. 2023:
- Feng Z, Zhu S, Duan J, Lu Y, Li L. Cross-modality effect in implicit learning of temporal sequence. *Curr Psychol*. 2023:
- Friedrich J, Verrel J, Kleimaker M, Münchau A, Beste C, Bäumer T. Neurophysiological correlates of perception–action binding in the somatosensory system. *Sci Rep*. 2020;10:14794.
- Frings C, Hommel B, Koch I, Rothermund K, Dignath D, Giesen C, Kiesel A, Kunde W, Mayr S, Moeller B, et al. Binding and retrieval in action control (BRAC). *Trends Cogn Sci*. 2020;24(5):375–387.
- Frost R, Armstrong BC, Siegelman N, Christiansen MH. Domain generality versus modality specificity: the paradox of statistical learning. *Trends Cogn Sci*. 2015;19:117–125.
- Goschke T. Implicit learning of perceptual and motor sequences: evidence for independent learning systems. In: *Handbook of implicit learning*. Thousand Oaks (CA): Sage Publications, Inc.; 1998. pp. 401–444.
- Gottlieb J. From thought to action: the parietal cortex as a bridge between perception, action, and cognition. *Neuron*. 2007;53:9–16.
- Gottlieb J, Snyder LH. Spatial and non-spatial functions of the parietal cortex. *Curr Opin Neurobiol Motor Syst Neurobiol Behav*. 2010;20:731–740.
- Haider H, Esser S, Eberhardt K. Feature codes in implicit sequence learning: perceived stimulus locations transfer to motor response locations. *Psychol Res*. 2020;84:192–203.
- Hallgató E, Györi-Dani D, Pekár J, Janacsek K, Nemeth D. The differential consolidation of perceptual and motor learning in skill acquisition. *Cortex*. 2013;49:1073–1081.
- Han YC, Reber PJ. Implicit sequence learning using auditory cues leads to modality-specific representations. *Psychon Bull Rev*. 2022;29(2):541–551.
- Hommel B. Event files: feature binding in and across perception and action. *Trends Cogn Sci*. 2004;8:494–500.
- Hommel B. Theory of event coding (TEC) V2.0: representing and controlling perception and action. *Atten Percept Psychophys*. 2019;81:2139–2154.
- Hommel B, Frings C. The disintegration of event files over time: Decay or interference? *Psychon Bull Rev*. 2020;27:751–757.
- Hommel B, Müsseler J, Aschersleben G, Prinz W. The theory of event coding (TEC): a framework for perception and action planning. *Behav Brain Sci*. 2001;24:849–878.
- Horváth K, Nemeth D, Janacsek K. Inhibitory control hinders habit change. *Sci Rep*. 2022;12:8338.
- Howard JH, Howard DV. Age differences in implicit learning of higher order dependencies in serial patterns. *Psychol Aging*. 1997;12:634–656.
- Janacsek K, Ambrus GG, Paulus W, Antal A, Nemeth D. Right hemisphere advantage in statistical learning: evidence from a probabilistic sequence learning task. *Brain Stimulat*. 2015;8:277–282.
- Janacsek K, Shattuck KF, Tagarelli KM, Lum JAG, Turkeltaub PE, Ullman MT. Sequence learning in the human brain: a functional neuroanatomical meta-analysis of serial reaction time studies. *NeuroImage*. 2020;207:116387.

- Janacek K, Evans TM, Kiss M, Shah L, Blumenfeld H, Ullman MT. Subcortical cognition: the fruit below the rind. *Annu Rev Neurosci*. 2022;45(1):361–386.
- Kanwisher N, Yovel G. The fusiform face area: a cortical region specialized for the perception of faces. *Philos Trans R Soc B Biol Sci*. 2006;361:2109–2128.
- Kayser J, Tenke CE. On the benefits of using surface Laplacian (current source density) methodology in electrophysiology. *Int J Psychophysiol Off J Int Organ Psychophysiol*. 2015;97:171–173.
- Kemény F, Lukács Á. Sequence in a sequence: learning of auditory but not visual patterns within a multimodal sequence. *Acta Psychol*. 2019;199:102905.
- Kóbor A, Takács Á, Kardos Z, Janacek K, Horváth K, Csépe V, Nemeth D. ERPs differentiate the sensitivity to statistical probabilities and the learning of sequential structures during procedural learning. *Biol Psychol*. 2018;135:180–193.
- Kóbor A, Horváth K, Kardos Z, Takács Á, Janacek K, Csépe V, Nemeth D. Tracking the implicit acquisition of nonadjacent transitional probabilities by ERPs. *Mem Cogn*. 2019;47:1546–1566.
- Lukács Á, Kemény F. Development of different forms of skill learning throughout the lifespan. *Cogn Sci*. 2015;39:383–404.
- Lum JAG, Clark GM, Barhoun P, Hill AT, Hyde C, Wilson PH. Neural basis of implicit motor sequence learning: modulation of cortical power. *Psychophysiology*. 2023;60:e14179.
- Maheu M, Meyniel F, Dehaene S. Rational arbitration between statistics and rules in human sequence processing. *Nat Hum Behav*. 2022;6:1087–1103.
- Marco-Pallarés J, Grau C, Ruffini G. Combined ICA-LORETA analysis of mismatch negativity. *NeuroImage*. 2005;25:471–477.
- Mückschel M, Chmielewski W, Ziemssen T, Beste C. The norepinephrine system shows information-content specific properties during cognitive control—evidence from EEG and pupillary responses. *NeuroImage*. 2017;149:44–52.
- Mullen T, Kothe C, Chi YM, Ojeda A, Kerth T, Makeig S, Cauwenberghs G, Jung T-P. Real-time modeling and 3D visualization of source dynamics and connectivity using wearable EEG. *Annu Int Conf IEEE Eng Med Biol Soc IEEE Eng Med Biol Soc Annu Int Conf*. 2013;2013:2184–2187.
- Nemeth D, Hallgató E, Janacek K, Sándor T, Londe Z. Perceptual and motor factors of implicit skill learning. *Neuroreport*. 2009;20:1654–1658.
- Nemeth D, Janacek K, Fiser J. Age-dependent and coordinated shift in performance between implicit and explicit skill learning. *Front Comput Neurosci*. 2013;7:
- Nemeth D, Janacek K, Polner B, Kovacs ZA. Boosting human learning by hypnosis. *Cereb Cortex*. 2013;23:801–805.
- Ocklenburg S, Friedrich P, Fraenz C, Schlüter C, Beste C, Güntürkün O, Genç E. Neurite architecture of the planum temporale predicts neurophysiological processing of auditory speech. *Sci Adv*. 2018;4:eaar6830.
- Opitz A, Beste C, Stock A-K. Using temporal EEG signal decomposition to identify specific neurophysiological correlates of distractor-response bindings proposed by the theory of event coding. *NeuroImage*. 2020;209:116524.
- Ouyang G, Zhou C. Characterizing the brain's dynamical response from scalp-level neural electrical signals: a review of methodology development. *Cogn Neurodyn*. 2020;14(6):731–742.
- Ouyang G, Herzmann G, Zhou C, Sommer W. Residue iteration decomposition (RIDE): a new method to separate ERP components on the basis of latency variability in single trials. *Psychophysiology*. 2011;48:1631–1647.
- Ouyang G, Sommer W, Zhou C. A toolbox for residue iteration decomposition (RIDE)—a method for the decomposition, reconstruction, and single trial analysis of event related potentials. *J Neurosci Methods*. Cutting-edge EEG Methods. 2015a;250:7–21.
- Ouyang G, Sommer W, Zhou C. Updating and validating a new framework for restoring and analyzing latency-variable ERP components from single trials with residue iteration decomposition (RIDE): ERP analysis with residue iteration decomposition. *Psychophysiology*. 2015b;52:839–856.
- Park J, Janacek K, Nemeth D, Jeon H-A. Reduced functional connectivity supports statistical learning of temporally distributed regularities: abbreviated title: functional connectivity in statistical learning. *NeuroImage*. 2022;260:119459.
- Parra LC, Spence CD, Gerson AD, Sajda P. Recipes for the linear analysis of EEG. *NeuroImage*. 2005;28:326–341.
- Pascual-Marqui RD, Esslen M, Kochi K, Lehmann D. Functional imaging with low-resolution brain electromagnetic tomography (LORETA): a review. *Methods Find Exp Clin Pharmacol*. 2002;24:91–95.
- Pedroni A, Bahreini A, Langer N. Automagic: Standardized preprocessing of big EEG data. *NeuroImage*. 2019;200:460–473.
- Perrin F, Pernier J, Bertrand O, Echallier JF. Spherical splines for scalp potential and current density mapping. *Electroencephalogr Clin Neurophysiol*. 1989;72:184–187.
- Petruo V, Takacs A, Mückschel M, Hommel B, Beste C. Multi-level decoding of task sets in neurophysiological data during cognitive flexibility. *iScience*. 2021;24:103502.
- Pion-Tonachini L, Kreutz-Delgado K, Makeig S. The ICLabel dataset of electroencephalographic (EEG) independent component (IC) features. *Data Brief*. 2019;25:104101.
- Prochnow A, Bluschke A, Weissbach A, Münchau A, Roessner V, Mückschel M, Beste C. Neural dynamics of stimulus-response representations during inhibitory control. *J Neurophysiol*. 2021;126:680–692.
- Remillard G. Pure perceptual-based sequence learning. *J Exp Psychol Learn Mem Cogn*. 2003;29:581–597.
- Rüsseler J, Rösler F. Implicit and explicit learning of event sequences: evidence for distinct coding of perceptual and motor representations. *Acta Psychol*. 2000;104:45–67.
- Schreier ML, Chmielewski W, Beste C. Neurophysiological processes and functional neuroanatomical structures underlying proactive effects of emotional conflicts. *NeuroImage*. 2018;174:11–21.
- Sekihara K, Sahani M, Nagarajan SS. Localization bias and spatial resolution of adaptive and non-adaptive spatial filters for MEG source reconstruction. *NeuroImage*. 2005;25:1056–1067.
- Song S, Howard JH, Howard DV. Perceptual sequence learning in a serial reaction time task. *Exp Brain Res*. 2008;189:145–158.
- Southwell R, Chait M. Enhanced deviant responses in patterned relative to random sound sequences. *Cortex*. 2018;109:92–103.
- Stark-Inbar A, Raza M, Taylor JA, Ivry RB. Individual differences in implicit motor learning: task specificity in sensorimotor adaptation and sequence learning. *J Neurophysiol*. 2017;117:412–428.
- Stock A-K, Arming L, Epplen JT, Beste C. DRD1 and DRD2 genotypes modulate processing modes of goal activation processes during action cascading. *J Neurosci*. 2014;34:5335–5341.
- Stürmer B, Ouyang G, Zhou C, Boldt A, Sommer W. Separating stimulus-driven and response-related LRP components with residue iteration decomposition (RIDE). *Psychophysiology*. 2013;50:70–73.
- Takács Á, Kóbor A, Chezan J, Éltető N, Tárnok Z, Nemeth D, Ullman MT, Janacek K. Is procedural memory enhanced in Tourette syndrome? Evidence from a sequence learning task. *Cortex*. Embodiment disrupted: Tapping into movement disorders through syntax and action semantics. 2018;100:84–94.

- Takacs A, Mückschel M, Roessner V, Beste C. Decoding stimulus-response representations and their stability using EEG-based multivariate pattern analysis. *Cereb Cortex Commun.* 2020;1(1).
- Takacs A, Zink N, Wolff N, Münchau A, Mückschel M, Beste C. Connecting EEG signal decomposition and response selection processes using the theory of event coding framework. *Hum Brain Mapp.* 2020;41:2862–2877.
- Takacs A, Bluschke A, Kleimaker M, Münchau A, Beste C. Neurophysiological mechanisms underlying motor feature binding processes and representations. *Hum Brain Mapp.* 2021;42:1313–1327.
- Takács Á, Kóbor A, Kardos Z, Janacsek K, Horváth K, Beste C, Nemeth D. Neurophysiological and functional neuroanatomical coding of statistical and deterministic rule information during sequence learning. *Hum Brain Mapp.* 2021;42:3182–3201.
- Takacs A, Münchau A, Nemeth D, Roessner V, Beste C. Lower-level associations in Gilles de la Tourette syndrome: convergence between hyperbinding of stimulus and response features and procedural hyperfunctioning theories. *Eur J Neurosci.* 2021;54:5143–5160.
- Takács Á, Yu S, Mückschel M, Beste C. Protocol to decode representations from EEG data with intermixed signals using temporal signal decomposition and multivariate pattern-analysis. *STAR Protoc.* 2022;3:101399.
- Thiessen ED, Erickson LC. Beyond word segmentation: a two-process account of statistical learning. *Curr Dir Psychol Sci.* 2013;22:239–243.
- Tóth B, Janacsek K, Takács Á, Kóbor A, Zavec Z, Nemeth D. Dynamics of EEG functional connectivity during statistical learning. *Neurobiol Learn Mem.* 2017;144:216–229.
- Treder MS. MVPA-light: a classification and regression toolbox for multi-dimensional data. *Front Neurosci.* 2020:14.
- Vahid A, Stock A-K, Mückschel M, Beste C. On the relative importance of attention and response selection processes for multi-component behavior—evidence from EEG-based deep learning. *Neuroimage Rep.* 2022;2:100118.
- Verleger R, Metzner MF, Ouyang G, Śmigajewicz K, Zhou C. Testing the stimulus-to-response bridging function of the oddball-P3 by delayed response signals and residue iteration decomposition (RIDE). *Neuroimage.* 2014;100:271–280.
- Widmann A, Schröger E, Maess B. Digital filter design for electrophysiological data—a practical approach. *J Neurosci Methods.* 2015;250:34–46.
- Winkler I, Haufe S, Tangermann M. Automatic classification of artifactual ICA-components for artifact removal in EEG signals. *Behav Brain Funct.* 2011;7:30.
- Winkler I, Brandl S, Horn F, Waldburger E, Allefeld C, Tangermann M. Robust artifactual independent component classification for BCI practitioners. *J Neural Eng.* 2014;11:035013.
- Wolff N, Mückschel M, Beste C. Neural mechanisms and functional neuroanatomical networks during memory and cue-based task switching as revealed by residue iteration decomposition (RIDE) based source localization. *Brain Struct Funct.* 2017;222:3819–3831.
- Yu S, Mückschel M, Hoffmann S, Bluschke A, Pscherer C, Beste C. The neural stability of perception–motor representations affects action outcomes and behavioral adaptation. *Psychophysiology.* 2023;60(1):e14146.