OXFORD

Cerebral Cortex, 2025, **35**, bhaf025
https://doi.org/10.1093/cercor/bhaf025
Advance access publication date 24 February 2025
Original Article

Sequence-dependent predictive coding during the learning and rewiring of skills

Ádám Takács[1,2,†], Teodóra Vékony[3,4,†,*], Felipe Pedraza[4,5], Frederic Haesebaert[6], Barbara Tillmann[7], Christian Beste[1,2], and Dezső Németh[3,4,8]

[1]Cognitive Neurophysiology, Department of Child and Adolescent Psychiatry, Faculty of Medicine, TU Dresden, Fetscherstraße, Fetscherstrasse 74, 01309, Dresden, Germany
[2]University Neuropsychology Center Faculty of Medicine, TU Dresden, Fetscherstrasse 74, 01309, Dresden, Germany
[3]Gran Canaria Cognitive Research Center, Department of Education and Psychology, University of Atlántico Medio, Ctra. de Quilmes, 37, 35017, Tafira Baja, Las Palmas de Gran Canaria, Spain
[4]Centre de Recherche en Neurosciences de Lyon, INSERM, CNRS, Université Claude Bernard Lyon 1 CRNL, 95 Bd Pinel, 69500, Bron, France
[5]Laboratory EMC (EA 3082), Université de Lyon Université Lyon 2, 5 Av. Pierre Mendès France, 69500, Bron, France
[6]Centre de Recherche en Neurosciences de Lyon, INSERM, CNRS, Université Claude Bernard Lyon 1 CRNL U1028 UMR5292, PSYR2 Team, 95 Bd Pinel, 69005, Bron, France
[7]CNRS, UMR5022, Laboratoire d'Etude de l'Apprentissage et du Développement, Université Bourgogne Europe, 11 Esplanade Erasme, 21000, Dijon, France
[8]BML-NAP Research Group, Institute of Psychology Eötvös Loránd University & Institute of Cognitive Neuroscience and Psychology, Hun-Ren Research Centre for Natural Sciences, Damjanich utca 41, 1072, Budapest, Hungary

*Corresponding author: Teodóra Vékony, Department of Education and Psychology, University of Atlántico Medio, Ctra. de Quilmes, 37, 35017 Tafira Baja, Las Palmas, Spain. Email: teodora.vekony@pdi.atlanticomedio.es

†Ádám Takacs and Teodóra Vékony contributed equally to this work.

In the constantly changing environment that characterizes our daily lives, the ability to predict and adapt to new circumstances is crucial. This study examines the influence of sequence and knowledge adaptiveness on predictive coding in skill learning and rewiring. Participants were exposed to two different visuomotor sequences with overlapping probabilities. By applying temporal decomposition and multivariate pattern analysis, we dissected the neural underpinnings across different levels of signal coding. The study provides neurophysiological evidence for the influence of knowledge adaptiveness on shaping predictive coding, revealing that these are intricately linked and predominantly manifest at the abstract and motor coding levels. These findings challenge the traditional view of a competitive relationship between learning context and knowledge, suggesting instead a hierarchical integration where their properties are processed simultaneously. This integration facilitates the adaptive reuse of existing knowledge in the face of new learning. By shedding light on the mechanisms of predictive coding in visuomotor sequences, this research contributes to a deeper understanding of how the brain navigates and adapts to environmental changes, offering insights into the foundational processes that underlie learning and adaptation in dynamic contexts.

Keywords: electrophysiology; multivariate pattern analysis; statistical learning; skill learning.

Introduction

In volatile environments, temporal predictions can help us use our resources more efficiently. For instance, statistical learning, a key component in skill learning, allows us to differentiate between probable and less probable events. This ability leads to more optimal processing both at the behavioral (Fiser and Aslin 2001; Aslin 2017) and neural levels (Turk-Browne et al. 2009; Takacs and Beste 2023), ultimately facilitating the acquisition and refinement of new skills. Exposure to repeated temporal associations creates robust long-term memory trace (Kobor et al. 2017; Tóth-Fáber et al. 2021). That is, existing knowledge with predictive associations can be observed as a neural representation (Vékony et al. 2023). However, the effectiveness of learned predictions varies across sequences. When environments change, previously learned temporal associations may no longer be accurate, requiring adaptation to maintain predictive accuracy (ie to rewire old skills) (Heald et al. 2023a; Heald et al. 2023b) (Fig. 1). Yet, we have limited knowledge of how contextual factors and knowledge effects contribute to predictions (Apps and Tsakiris 2013), especially in a volatile environment. We aimed to fill this gap by investigating how previously learned associations are utilized across different sequences, each varying in the extent to which knowledge (old or new) can be effectively applied. Essentially, we wanted to understand the rewiring process, that is, how the brain accommodates new information while retaining previously learned associations, depending on the specific learning sequence. We examined this process at different coding levels of the neurophysiological signal to gain a comprehensive understanding of the underlying mechanisms.

In everyday life, familiar environments, such as supermarkets, restaurants, and bus stops, are already learnt and useful for making predictions over extended periods. Nevertheless, the volatility of these environments can limit its predictive value, as exemplified by shopping in a familiar store that just had its aisles reshuffled. The dual-process theory of memory retrieval (Yonelinas 1994) proposed that this is an automatic, context-free process as opposed to conscious recollection that comes with the reactivation of the episodic details. However, neurophysiological

Received: June 17, 2024. **Revised:** December 3, 2024. **Accepted:** January 23, 2025

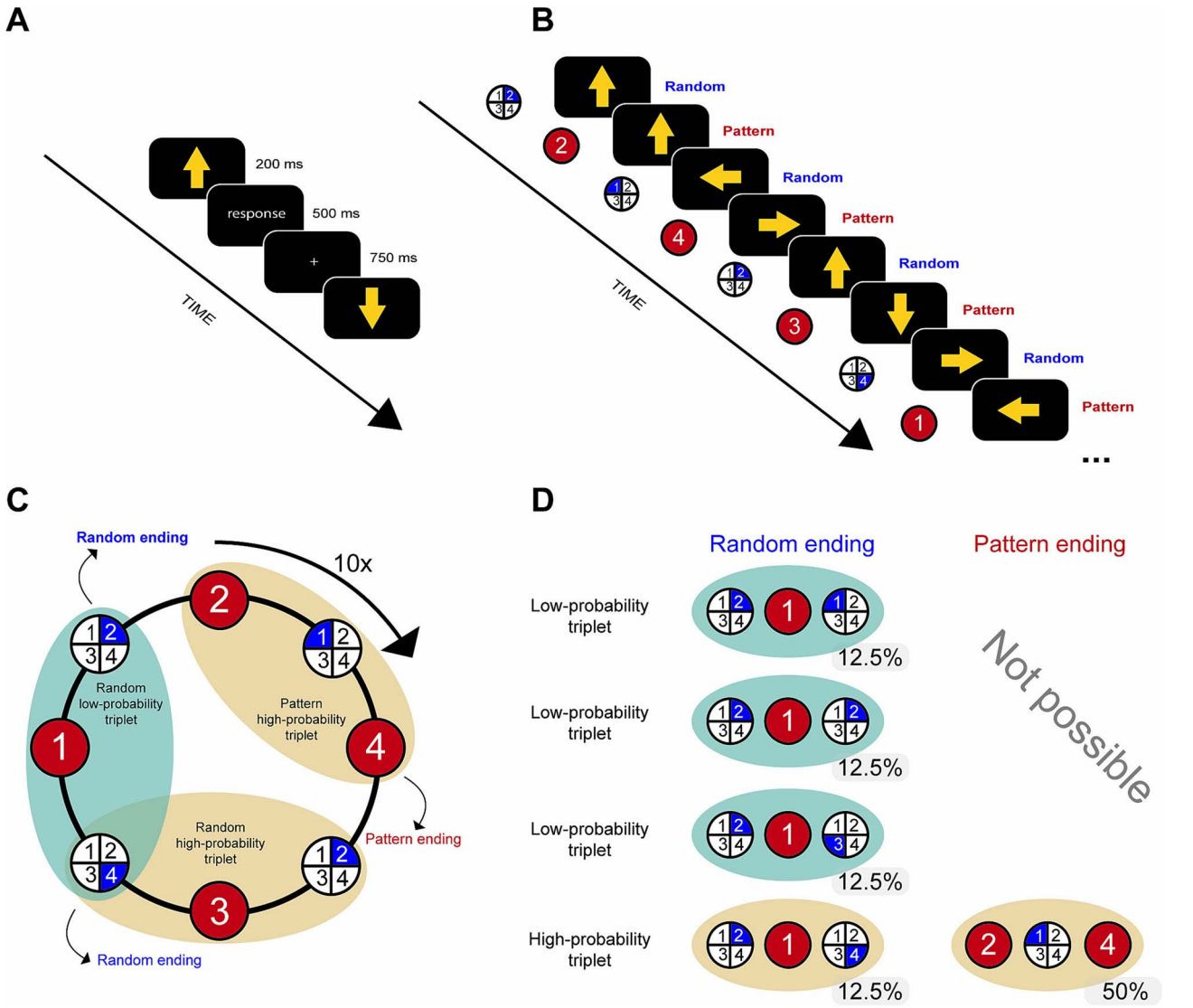

Fig. 1. Experimental task. A) The timing of the ASRT task. One arrow pointing left, up, down, or right was presented in the middle of the screen. B) The structure of the ASRT task. Unbeknownst to the participants, the order of the direction of the arrows followed a predetermined, alternating order. An 8-element alternating sequence was repeated 10 times per block. Every second element was a pattern trial, while the intermittent elements appeared in a randomly chosen position. C) The formation of triplets. Due to the alternating structure, three consecutive elements formed a high- or low-probability triplet. We categorized each trial as the last element of a high- or low-probability triplet. D) The proportion of high- and low-probability triplets. Thirty-seven point five percent of the triplets were low probability, while 62.5% of the trials were high probability. High-probability trials could have been formed by two pattern trials and one random trial in the middle (50% of the time) or, by chance, also by two random trials and one pattern trial in the middle.

markers of knowledge shown to be modulated by the environment (Tsivilis et al. 2001). An alternative account suggested (Ecker et al. 2007) that although information about the environment is predominantly processed as noise in decisions if the context is salient enough, it can function as an independent familiarity signal. As a result, context and knowledge compete for attention. In contrast, predictive coding accounts suggest that contextual and stimulus properties contribute to the evaluation of the perceived stream of information in a simultaneous fashion (Apps and Tsakiris 2013; Blank et al. 2023). Specifically, contexts' structure and characteristics can shape expectations about probable stimulus properties, which can bias the stimulus processing. In the case of sequential regularities, this implies that the given context (sequence) influences the recognition of which contingencies are more probable to occur.

Irrespective of whether the learning context and knowledge are competing (Ecker et al. 2007), processing them simultaneously presumes intermixed neurophysiological signals (Mückschel et al. 2017; Takács et al. 2021). Temporal signal decomposition can be used to disentangle these partially overlapping processes (Ouyang and Zhou 2020; Takács et al. 2022). This has also been shown in learning visuomotor sequences (Takács et al. 2021; Vékony et al. 2023). In a recent study (Vékony et al. 2023), we employed Residue Iteration Decomposition (RIDE) to show that perceptual (S-cluster), motor (R-cluster), and abstract or modality-independent (C-cluster) information simultaneously contribute to the formation of memory representations in an environment that contained sequential regularities.

We aimed to study the effects of sequence (whether sequence A or B is presented) and knowledge of predictive associations (old and new knowledge) in learned but volatile environments that mimic these everyday challenges. Therefore, we implemented a method from the habit change literature (Szegedi-Hallgató et al. 2017; Horváth et al. 2022). Using a visuomotor sequence

learning paradigm, participants were exposed to different sequences that either (i) they had previously encountered, allowing them to develop implicit knowledge of adaptive responses, or (ii) were entirely new to them, with no prior experience. However, some of the responses learned from sequence A remained effective and adaptive when participants were exposed to sequence B. This approach enabled us to examine the persistence of old knowledge from the original sequence (ie exposure to sequence A) and how new adaptive knowledge emerges in response to changes in the environment (ie exposure to sequence B).

In the current study, we investigated the development of the initially formed associations (old knowledge) under a new sequence that required the learning of novel associations. Our experimental design is unique in that it allows us to examine the impact of sequence—that is, in our case, the specific probabilistic sequence currently presented to the participant—and determine whether the knowledge being activated pertains to a previously learned sequence (first session) or a newly learned sequence (second session). Given that certain sequence items (triplets) are acquired within one sequence while others are learned in a different one, we distinguish between them by referring to the former as "*Old Knowledge*" and the latter as "*New Knowledge*." After the second session, we also tested the participants' knowledge as they experienced alternating sequences (ie whether the old or new knowledge is advantageous in the given sequence, third session). Importantly, participants were unaware of changes between sequences. Moreover, the learned regularities of the sequences partially overlapped; therefore, switching from the old to the new sequence required a gradual adaptation.

Previous studies on habit change suggest that it is possible to acquire new habits without losing the old ones (Apps and Tsakiris 2013; Horváth et al. 2022; Székely et al. 2024). Based on these findings, we hypothesized that participants could learn in the new regularities in both sequence A and then sequence B while retaining their knowledge of Sequence A. We further predicted that stimulus contingencies could be decoded at all levels of processing, including the sensory (S-cluster), motor (R-cluster), and cognitive (C-cluster) coding levels (Vékony et al. 2023). However, we anticipated that sequence-specific modulations would be observed at the perceptual (Tsivilis et al. 2001; Apps and Tsakiris 2013) and motor coding levels.

Materials and methods

Participants

Forty-four healthy young participants were recruited via online advertisements. Participants were right-handed and had normal or corrected vision. They had no history of neurological or psychiatric conditions and were not taking psychoactive medications at the time of the study. Three participants were excluded because they did not complete data acquisition in all three sessions. One participant was excluded due to low signal quality on the first day (<10 trials remaining), leaving 40 participants in the final sample (19 female, 21 male, $M_{age} = 22.6$ years, SD = 2.9). They all signed a written consent form before taking part and received 20€/h as compensation. The study followed the declaration of Helsinki and was approved by the relevant ethical committee ("Comité de Protection des Personnes Est I," ID RCB 2019-A02510-57).

Task

The electroencephalography (EEG)-adapted version of the Alternating Serial Reaction Time (ASRT) task (Kóbor et al. 2019; Vékony et al. 2023) was used to measure the learning of probabilistic regularities (Fig. 2). During this task, a yellow arrow appeared for 200 ms in the middle of a black screen, pointing left, up, down, or right. Then, a white cross was displayed for 500 ms. Participants had to press a button that matched the arrow's direction during this time. If they did, the cross stayed for another 750 ms. Erroneous or missing responses were followed by an "X" or an "!," respectively, for 500 ms. This was followed by the presentation of the fixation cross for 250 ms.

The task had blocks of 85 trials each. Blocks started with five warm-up trials (random arrows), and then an eight-element sequence was reiterated 10 times. Participants were not informed that the arrows followed a fixed alternating sequence, and they were not briefed about the purpose of the task between sessions. In the sequence of eight elements, pattern (P) and pseudo-random (r) elements took turns (eg r-2-r-4-r-3-r-1, where numbers stand for a fixed spatial position in the sequence [1 = left, 2 = up, 3 = down, 4 = right], and r denotes a random position). There are 24 different ways to create an eight-element alternating sequence with four spatial directions (eg r-2–r–1–r–3–r–4; r-1–r–2–r–3–r–4; r-1–r–2–r–4–r–3). Within these permutations, only six are unique (eg r-2–r–4–r–3–r–1 is the same as r-3–r–1–r–2–r–4).

Each trial of the ASRT task can be categorized as the third element of a triplet based on the preceding two trials (in either rPr or PrP structures). Importantly, due to the alternating sequence, we can differentiate between high-probability and low-probability triplets based on how predictable the last (third) element of a triplet is from the first element: the third element of a high-probability triplet can be predicted from the first element with a higher chance than in a low-probability triplet (second-order transitional probabilities). Because of the eight-element sequence, 64 different triplets can be formed. If we know the first two elements of the triplet (eg 2-1), there are four possible endings for the third element: The triplet could end with 1, 2, 3, or 4. Only one of them will make a high-probability triplet: using the example above (r-2-r-4-r-3-r-1). That is, only "4" as the third element will make a high-probability triplet (in that case, the triplet will be 2-1-4). Consequently, out of the 64 possible triplets, 16 will be high-probability and 48 will be low-probability triplets. Half of the trials will be the last element of a high-probability triplet (because every other element is part of the pattern, P-r-P structure). However, one out of four triplets with an r-P-r structure will also be the last element of a high-probability triplet (50%/4 = 12.5%). This is because 2-X-4, which represents a high-probability triplet based on the example provided, could be formed from a PrP structure but occasionally also from an rPr structure. Thus, a specific high-probability triplet will show up five times more often (50% + 12.5% = 62.5% of the trials) than a low-probability triplet (37.5% of the trials).

Participants received different sequences of stimulus presentation in the first two sessions (sequence A and sequence B, respectively). In the third session, sequences A and B alternated in runs of 5 blocks. The order of the sequence presentation was counterbalanced among participants who performed them either as A-B-A-B-A-B or as B-A-B-A-B-A. The first two sessions encompassed 25 blocks each, while the third session comprised 30 blocks. Reaction times (RTs) and accuracy were recorded. After each block, participants received feedback on their performance, displayed as an average RT and accuracy, irrespective of experimental conditions. This feedback remained on the screen for 5,000 ms, succeeded by a 15,000-ms break, which participants could prolong if needed. The commencement of a new block was self-paced.

Procedure

Participants visited the lab on three consecutive days to complete the three sessions. Each session followed a similar procedure.

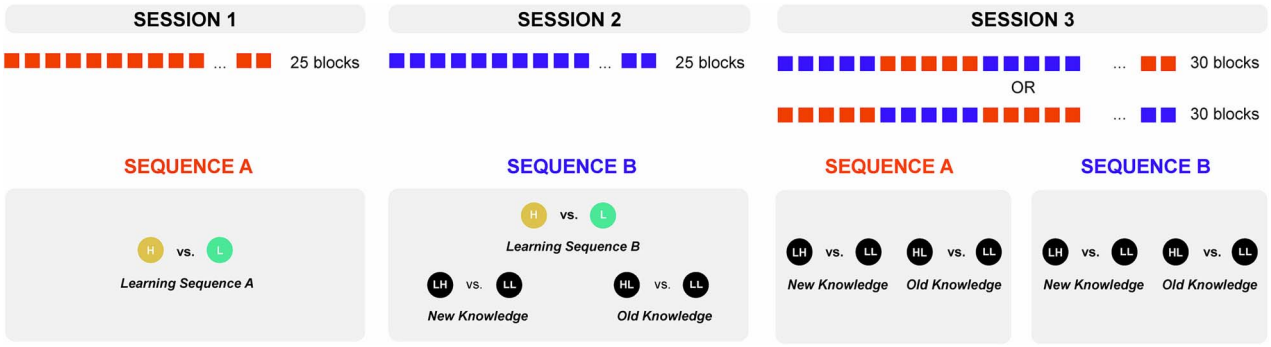

Fig. 2. Study design. The study consisted of three consecutive sessions. In session 1, participants learned sequence A, and, in session 2, they learned sequence B, each presented in 25 blocks of trials. Learning was assessed by comparing performance on high- vs. low-probability trials in both sessions. In session 3, participants completed 15 blocks each of sequence A and sequence B, with the starting sequence counterbalanced. Sequences were presented in units of five consecutive blocks. To assess knowledge updating in sessions 2 and 3, we compared performance on trials that were low probability on the first day but became high probability on the second day ("New Knowledge") to trials that remained low probability throughout. To assess knowledge retention in sessions 2 and 3, we compared performance on trials that were high probability on the first day but became low probability on the second day ("Old Knowledge") to trials that remained low probability throughout.

The experiment began with a 5-min eyes-open resting-state EEG recording, which was not analyzed for the current paper. During this period, participants looked at a white cross on a black background without moving. Next, they performed the ASRT task (Howard and Howard 1997; Kóbor et al. 2019). In this task, yellow arrows sequentially appeared in the center of the screen, and participants responded by pressing a corresponding button on a Cedrus RB-530 response pad (Cedrus Corporation, San Pedro, CA). Participants were instructed to press the button corresponding to the direction of the arrow as quickly and accurately as possible. They kept their left and right thumbs and left and right index fingers on the four buttons. Initially, they completed three blocks of 85 stimuli each, where the arrows pointed in random directions to get used to the setting (practice blocks). After that, they completed 25 blocks (85 stimuli/block) of the ASRT task (see details below). They rested for ~5 min after every five blocks, and EEG impedance was checked and adjusted if needed.

In the first session, participants implicitly learned to differentiate between high and low-probability trials (Vékony et al. 2023). We refer to this as "*Learning Sequence A*" (Fig. 2). Unbeknown to participants, the conditional probabilities of the stimulus sequence were changed in the second session. We refer to this part of the experiment as "*Learning Sequence B*." Finally, in the third session (testing), participants performed an extended version of the task (ie 30 blocks instead of 25). Half of the participants started the task with stimulus probabilities of sequence A, while the other half of the sample performed it with the probabilities from sequence B. After each five blocks, the sequences alternated. That is, every participant performed the task with stimulus probabilities from both sequences for 15 blocks each. This alternating presentation of sequences ensured counterbalancing to mitigate potential interference effects between old and new knowledge (Fig. 2).

Behavioral analyses

In the ASRT paradigm, probabilistic sequence learning is conventionally quantified as the difference between high- and low-probability trials (Howard and Howard 1997; Nemeth et al. 2013; Szegedi-Hallgató et al. 2019). As participants become familiar with the patterns, they exhibit shorter RTs and greater accuracy on high-probability trials compared to low-probability ones. This suggests that participants develop predictions based on the non-adjacent transitional probabilities, which subsequently leads to response adaptation.

In the current study, we only analyzed trials with correct responses. We also removed trills (eg 1-2-1, 8.73% of the trials) and repetitions (eg 1-1-1, 2.74% of the trials) from the analysis, as participants often have pre-existing biases toward these response configurations (Nemeth et al. 2013; Éltető et al. 2022). We also excluded the first seven trials, comprising the first five warm-up trials and the subsequent two, which constitute the first two elements of the first triplet (8.23% of the trials). The remaining trials were labeled as the last element of a high- or a low-probability triplet. For each participant and 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.

Additionally, trials were categorized based on the sequence (sequence A or sequence B). In the first session, participants performed sequence A (*Learning Sequence A*). The detailed results of this session, including learning curves, event-related potential (ERP) correlates, and decoding results, were reported in Vékony et al. (Vékony et al. 2023). In the second session, participants were exposed to Sequence B (*Learning Sequence B*). During this second session, the *old knowledge* was challenged by partially changed regularities. Participants were unaware of these changes, and the visual features of the experiment also remained unchanged. To acquire the new knowledge, participants needed to unlearn much of what they had learned in the first session (during learning sequence A), as it was no longer adaptive, while simultaneously learning new associations from the partially modified sequential regularities. Specifically, 75% of original high-probability triplets transitioned to low-probability status, while originally low-probability triplets became high-probability in sequence B. The difference between trials that remained low probability in both sequences and those that were originally high probability but became low probability in sequence B was quantified as *Old Knowledge* (Horváth et al. 2022). The difference between consistently low-probability trials and trials that were low probability in sequence A but became high probability in Sequence B was quantified as *New Knowledge*. Old and new knowledge was analyzed to quantify participants' reactions to volatile probabilities. Sensitivity to persistently adaptive knowledge was analyzed separately as unchanged probabilities: the difference between consistently low-probability trials and consistently high-probability trials (see Supplementary Results).

We also considered task reliability for the design of the experiment. Learning scores in the ASRT task are considered reliable in a neurotypical adult population with a sample size of $n = 21$ (Stark-Inbar et al. 2017). Also, learning scores in the ASRT task are reliable with 25 blocks of task length, based on internal consistency and split-half reliability measures (Farkas et al. 2024). Both sample size and task length met the reliability criteria.

EEG acquisition and preprocessing

We recorded EEG in a dimly lit room that was electrically and acoustically isolated. We used BrainAmp DC EEG amplifiers (BrainProducts GmbH, Gilching, Germany) to measure the signals from 64 actiCAP slim active electrodes on the scalp. The electrodes were placed according to the international 10–20 system in an elastic cap. The sampling rate was 500 Hz. We used AFz as the ground and the right side of the nose as a reference. We did not apply any online filters. The impedance of the electrodes was kept below 25 kΩ.

We used Automagic (Pedroni et al. 2019) and EEGLAB (Delorme and Makeig 2004) in Matlab 2019a (The MathWorks Corp.) to preprocess the EEG data. First, we removed any flat channels and re-referenced the data to an average reference. Then, we used the PREP preprocessing pipeline (Bigdely-Shamlo et al. 2015) to remove 50 Hz line noise with a multitaper algorithm and remove bad channels. Next, we used the *clean_rawdata* pipeline to (i) detrend the data using a finite impulse response (FIR) high-pass filter of 0.5 Hz (order 1286, stop-band attenuation 80 dB, transition band 0.25 to 0.75 Hz). In this step, we detected and removed flatline, noisy, and outlier channels. Then, we reconstructed time windows that had unusually high power (>15 SD compared to calibration data) using Artifact Subspace Reconstruction (ASR; burst criterion: 15) (Mullen et al. 2013). We removed any time windows that could not be reconstructed. We applied a low-pass filter of 40 Hz (sinc FIR filter; order: 86) (Widmann et al. 2015). We also removed electrooculography (EOG) artifacts using a subtraction method (Parra et al. 2005). We automatically identified and removed muscular and remaining eye-movement artifacts using an independent component analysis–based Multiple Artifact Rejection Algorithm (Winkler et al. 2011, 2014). We also identified and removed components that reflected cardiac artifacts using ICLabel (Pion-Tonachini et al. 2019). Finally, we interpolated all channels that were removed by Automagic using the spherical method. We segmented the preprocessed data separately for the three sessions. We created segments for high-probability and low-probability conditions for the whole task. In the second and third sessions, we segmented the data according to old knowledge and new knowledge, as well. Segments started −200 ms and ended 750 ms relative to stimulus onset. We only included segments with correct responses. We applied current-source density (CSD) transformation with four orders of splines to the segmented data (Perrin et al. 1989; Pion-Tonachini et al. 2019). CSD is a reference-free spatial filter that highlights scalp topography. Finally, we baseline-corrected the segmented data based on the 200-ms interval before the stimulus onset.

RIDE

We used RIDE (Ouyang et al. 2011, 2015a, 2015b; Ouyang and Zhou 2020) in Matlab 2019a (The MathWorks, Inc., Natick, Massachusetts, United States) as part of the RIDE-MVPA (multivariate pattern analyses) protocol, to decompose the neurophysiological signal based on temporal properties. 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 done on each channel separately. RIDE was used to estimate three clusters: The C-cluster ("central cluster") covers processes between stimulus and response, such as retrieval, decision-making, or response selection; the S-cluster ("stimulus cluster") collects information on stimulus-related processes, such as perception and attention; and the R-cluster ("response cluster") refers to motor preparation and execution (Ouyang and Zhou 2020). In each step of the iteration, the decomposition module estimates R, S, and C. For the estimation, RIDE subtracts the other two 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 used for further analysis. Several previous studies have shown that RIDE provides a conceptually meaningful separation of concomitant codings in the neurophysiological signal. The decomposition into three clusters requires predefined time windows for the initial cluster estimations. We have originally selected these time windows based on a previous study that used RIDE in EEG data with a modified ASRT paradigm (Takács et al. 2021): the time window between 300 ms before and 300 ms after the response markers for the R-cluster, 0 to 500 ms after stimulus onset for the S-cluster, and 150 to 600 ms after stimulus presentation for the C-cluster. In RIDE decomposition, an overlap between the initial search windows for RIDE clusters is a common practice (Wolff et al. 2017; Opitz et al. 2020; Ouyang and Zhou 2020), 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 validity of the three-cluster solution in the current paradigm was described in Vékony et al. (Vékony et al. 2023).

MVPA

We analyzed the single-trial data from the RIDE method using the MVPA-light toolbox (Treder 2020), in MATLAB (The MathWorks, Inc., Natick, Massachusetts, United States). We followed the protocol as described in Takács et al. The parameters were the same as in our previous study (Vékony et al. 2023); however, classes were defined differently. We decoded the difference between classes of "low-probability remained low-probability" and "high-probability became low-probability" as *Old knowledge*, and the difference between classes of "low-probability remained low-probability" and "low-probability became high-probability" as *New knowledge*. Each decoding was calculated for each RIDE cluster separately. Then, we ran temporal generalization analyses for the three class pairs in the three RIDE clusters to see how stable the decoded representations were over time at the different coding levels. Temporal generalization tests the classifier not only on the time points it was trained on (diagonal decoding) but also on the other time points. To avoid overfitting, we balanced the number of trials in each class by under-sampling. We used EEG amplitude data from 64 channels as features for classification. We chose an L1-Support Vector Machine (SVM) as a classifier with a cross-validation of five folds. In each case, we selected the area under the curve (AUC) as a measure of decoding performance and compared it to the chance level of AUC = 0.5 using Wilcoxon tests and cluster-based permutation.

Results

Behavioral results

First, we compared *Learning Sequence A* and *Learning Sequence B* (ie RT differences between high- and low-probability conditions separately in the two sequences). There was no significant difference between the learning of the two sequences: [$t(39) = 0.39$, $P = 0.700$, $d = 0.061$]. Second, we compared *Old Knowledge* and *New Knowledge* scores in the new context. There was no significant difference between them: [$t(39) = 1.57$, $P = 0.126$, $d = 0.247$]. Third, data from the third session was analyzed by using Sequence (sequence A vs. sequence B), Knowledge (Old Knowledge vs. New Knowledge), and Block Order (first, second, third instance) as factors in a repeated-measures ANOVA. No significant main effects were found for Sequence [$F(1, 39) = 1.98$, $P = 0.167$, $\eta p2 = 0.048$, $BF_{excl} = 0.948$], Knowledge [$F(1, 39) = 3.58$, $P = 0.066$, $\eta p2 = 0.084$, $BF_{excl} = 1.015$], or Block Order [$F(2, 78) = 0.80$, $P = 0.454$, $\eta p2 = 0.020$, $BF_{excl} = 14.961$]. However, the interaction between Sequence and Knowledge was significant [$F(1, 39) = 11.16$, $P = 0.002$, $\eta p2 = 0.222$, $BF_{excl} = 0.293$]. Bonferroni-corrected post hoc comparisons showed a larger expression of New Knowledge (5.87 ms ± 1.63) while performing sequence B (2.24 ms ± 1.46, $P = 0.002$). Similarly, the level of New Knowledge was greater during sequence B than during sequence A (1.65 ms ± 1.56, $P = 0.030$). Other pairwise differences were not significant (Ps > 0.239). Interactions between Sequence and Block Order [$F(2, 78) = 0.11$, $P = 0.899$, $\eta p2 = 0.003$, $BF_{excl} = 1857.596$[, Block Order and Knowledge [$F(2, 78) = 0.16$, $P = 0.856$, $\eta p2 = 0.004$, $BF_{excl} = 47.534$], and the three-way interaction were not significant [$F(2, 78) = 0.55$, $P = 0.580$, $\eta p2 = 0.014$, $BF_{excl} = 316.868$].

Thus, participants learned the predictabilities of the sequence B while keeping their Old Knowledge from sequence A. During testing, the New Knowledge was stronger than the Old Knowledge, and its strength changed according to its compatibility with the sequence.

Neurophysiological results

Figure 3A presents the decoding performance of *Old Knowledge* separately for the C-cluster, S-cluster, and R-cluster. Classification was significantly above chance in all three clusters, ranging from 90 to 609 ms after stimulus presentation in the S-cluster, from 168 to 625 ms in the C-cluster, and from 238 to 578 ms in the R-cluster. Similarly, Fig. 3A presents the decoding performance of *New Knowledge* separately for the three clusters. The classification was significantly above chance in all three coding levels, ranging from 86 to 609 ms after stimulus presentation in the S-cluster, from 160 to 652 ms in the C-cluster, and from 234 to 590 ms in the R-cluster.

In the S-cluster, decoding accuracy was the highest (ie AUC > 0.7) around the diagonal axis between 200 and 500 ms after stimulus presentations both in the *Old Knowledge* and the *New Knowledge* temporal generalization matrices (Fig. 3B). The accuracy decreased symmetrically from the axes with a jitter. Additionally, off-diagonal decoding patterns were observed: the time window of ∼150 to 400 ms of the unchanged probabilities generalized to the time window of 500 to 750 ms in the changed probabilities, respectively. Similarly, the time window of ∼100 to 350 ms of the changed probabilities is generalized to the time window of 500 to 750 ms in the unchanged probability class. In the C-cluster, the decoding accuracy peaked along the diagonal between 200 and 600 ms after stimulus onset (Fig. 3B). Accuracy decreased further from the axis in an asymmetrical fashion: In both generalization matrices, the cluster was larger above the diagonal than below. That is, representations of volatile probabilities ("high-probability became low-probability" and "low-probability became high-probability") generalized less to unchanged probabilities ("low-probability remained low-probability") than the other way around. In the R-cluster, decoding accuracy was the highest around the diagonal axis between 300 and 600 ms after stimulus presentations both in the *Old Knowledge* and the *New Knowledge* temporal generalization matrices (Fig. 3B). These loci decreased gradually toward the edge of the matrices.

Connecting behavioral and neurophysiological results

We conducted canonical correlation analyses to investigate the systematic relationship between sets of behavioral and neurophysiological variables. Specifically, we examined the relationship between learning old probabilities (ie *Learning Sequence A*; *Learning Sequence B*; *Old Knowledge*) and classification of *Old Knowledge* in the three coding levels. The full model was significant [$F(9, 82.90) = 2.19$, $\Lambda = 0.595$, $P = 0.031$, $Rc2 = 0.441$] with a canonical correlation coefficient of $r = 0.553$. Thus, better learning of the original probabilities was associated with better classification of *Old Knowledge*. The relevance of each variable in their respective latent measure is depicted in Fig. 4. Notably, *Learning Sequence A* had the largest canonical loading ($r = -0.850$) to the latent behavioral variable. Considering the latent neurophysiological variable, R-cluster ($r = 0.389$) and S-cluster ($r = 0.387$) loadings were similarly prominent. We also examined the relationship between learning new probabilities (ie *Learning Sequence B*; *New Knowledge*) and classification of *New Knowledge* in the three coding levels. The model was not significant [$F(6, 70) = 1.11$, $\Lambda = 0.834$, $P = 0.368$, $Rc2 = 0.159$, $r = 0.370$].

Next, we examined the relationship between the retained memory of old probabilities (ie *Old Knowledge* in sequence A and *Old knowledge* in sequence B) and the classification of Old Knowledge in the three coding levels. The model was significant [$F(6, 70) = 2.92$, $\Lambda = 0.465$, $P = 0.014$, $Rc2 = 0.465$] with a canonical correlation coefficient of $r = 0.563$. Thus, better retaining of the original probabilities was associated with better classification of *Old Knowledge*. Canonical loadings are presented in Fig. 4: The latent behavioral variable had the highest loading of *Old Knowledge* in *Sequence A* ($r = -0.995$), while the latent neurophysiological variable had the highest loading of the R-cluster classification ($r = -0.654$). Finally, we examined the relationship between the retained memory of new probabilities (ie *New Knowledge* in Sequence A and *New Knowledge* in sequence B) and the classification of *New Knowledge* in the three coding levels. The model was not significant [$F(6, 70) = 0.82$, $\Lambda = 0.873$, $P = 0.561$, $Rc2 = 0.138$, $r = 0.348$].

Discussion

Our study examined the development of temporal predictions as participants learned and adapted to a variety of sequences with different underlying structures. Participants successfully learned stimulus contingencies while performing both in the old and new sequences, even without knowing that the sequence had changed. While acquiring the new knowledge, participants also retained the original information. Thus, not only can old and new knowledge coexist, but both can also be retrieved when the context necessitates (Szegedi-Hallgató et al. 2017; Horváth et al. 2022). That is, sequence and knowledge adaptiveness both influenced behavioral performance. Moreover, decoding analyses highlighted the neurophysiological correlates of old and new knowledge while performing different sequences.

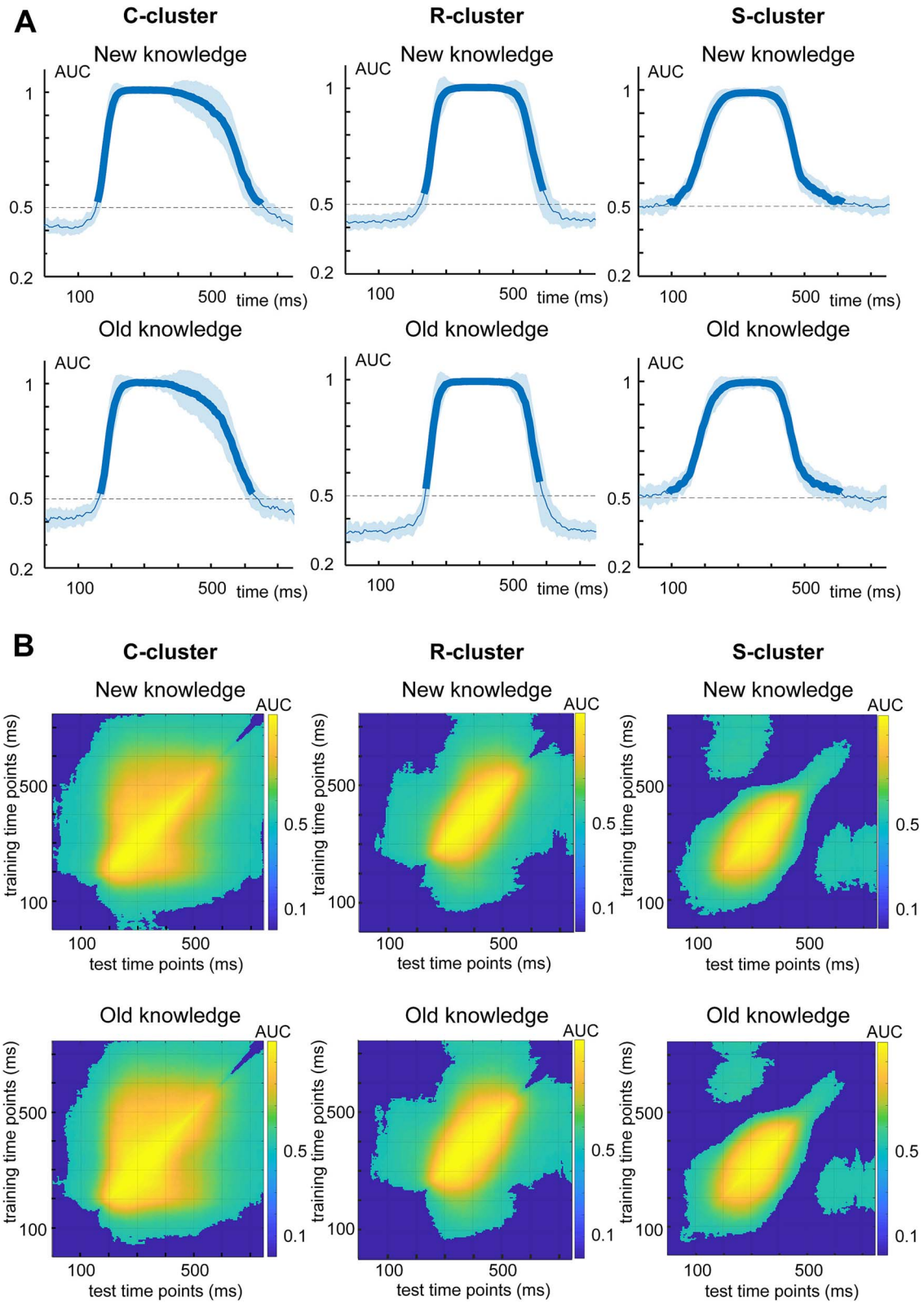

Fig. 3. Multivariate results: decoding and temporal generalization. A) Classification performance represented as AUC, separately for the RIDE decomposed C-, R-, and S-cluster EEG data. Time zero denotes the presentation of the target stimulus. Thicker lines indicate significant time windows ($P < 0.05$; two-sided cluster-based permutation). B) 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 colors indicate the classifier's performance.

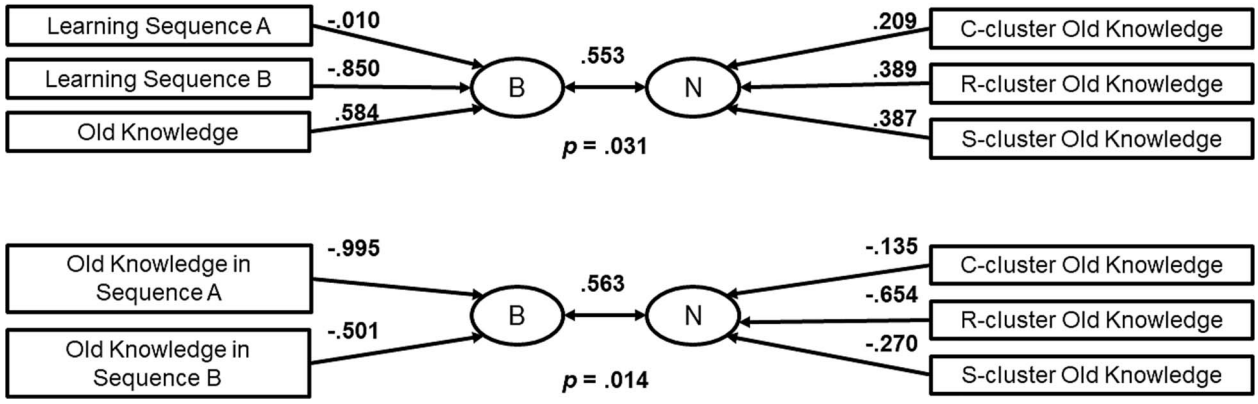

Fig. 4. Canonical correlations between behavioral and neurophysiological levels of analysis. The variables of the models are represented in rectangular shapes. Variables are connected to their respective synthetic variables by arrows. B denotes the behavioral, and N denotes the neurophysiological synthetic variable. The *r* values of the canonical loadings are presented above the arrows.

We expected to see the effect of knowledge adaptiveness and sequence in ERP mean amplitudes at the motor (Vékony et al. 2023) and perceptual (Ecker et al. 2007) coding levels. However, we only observed these effects at the motor level (see Supplementary Results and Supplementary Figure S1). The multivariate analyses, however, did not support the privileged role of motor coding. Both old and new knowledge were decoded on all three coding levels. The order of decoding onsets followed the pattern observed in the initial learning: first, the perceptual (S-cluster), followed by the modality-independent or abstract (C-cluster), and finally the motor level (R-cluster). This finding supports the view that old and new knowledge can coexist without interference between them (Horváth et al. 2022). Additionally, decoding accuracy and temporal generalization patterns were comparable between old and new knowledge on all three coding levels. Interestingly, while generalization was symmetrical between test and training classes for the S-cluster and R-cluster data, the results were asymmetrical in the C-cluster. Both old and new knowledge matrices indicated that the representation of unchanged probabilities generalized to volatile probabilities to a larger extent than the other way around. Thus, representations at the abstract coding level have the capacity to encode not only stimulus contingencies but also higher-order information, such as knowledge adaptiveness. A previous study emphasized the C-cluster's role in integrating statistical information into more abstract rules (Takács et al. 2021), a process that is closely tied to context learning (Heald et al. 2021; Székely et al. 2024).

Neither behavioral nor neurophysiological results support the competition between sequence and knowledge (Ecker et al. 2007). Old and new knowledge coexisted throughout the transition phase of the second sequence and both could be retrieved during testing. Thus, learning of new stimulus information does not appear to require either the extinction or the inhibition of previously acquired knowledge. However, the interaction between knowledge and context during the testing phase suggests that stimulus information is associated with the sequence. This is in line with the contextual inference (COIN) model (Heald et al. 2021; Heald et al. 2023a). The COIN presumes that continuous sensorimotor information is separated into a repertoire of memories. To determine whether it is sufficient to express or update an existing representation or if it is necessary to encode new memory, the learner must infer not only the general sensorimotor contingencies but also the context probability. If prediction error occurs only at the level of context (the sequence that is performed), it is more ecological to adjust the likelihood of the expression of different memories, rather than updating the memory traces. That is, sequence information is akin to rules or general task knowledge, and its relevance is primarily to aid the selection of the right action from the repertoire. Flexible action selection requires a two-level computation (Maheu et al. 2022; Székely et al. 2024), in which the lower level's parameter space represents regularities in the environment and a higher-order level represents the relevant structure. In the current study, amplitude modulations of the R-P2 component might reflect fine-tuning the parameter space for response selection, while the C-cluster decoding represents the sequence-dependent encoding of contingencies in the environment.

The canonical correlations between behavioral and neurophysiological levels of analyses also support the view that computations on different coding levels all contribute to the development and selection of internal models (Takács et al. 2021; Vékony et al. 2023). As measured by learning performance during the training instances, old knowledge was associated with decodings on all three levels. Additionally, the C-cluster, R-cluster, and S-cluster decoding accuracies of the old knowledge had comparable weights within the neurophysiological level. Interestingly, old knowledge in the testing session was primarily related to R-cluster decoding in the new context. It was suggested that motor code transfers more easily after training than the perceptual one (Hallgató et al. 2013), providing an advantage for consolidation. However, it is conceivable that decoding based on R-cluster data fits RT-based behavioral data more likely in a linear fashion than perceptual or abstract-level decodings could (Petruo et al. 2021; Vékony et al. 2023). Another reason for caution is that only old but not new contingencies were significantly correlated with the neurophysiological levels. Old knowledge was originally acquired in sequence A, accessed during the training in sequence B, and expressed during testing in both sequences. The lack of significant correlations with the new knowledge might be an artifact of a shorter exposure than with the old knowledge. Alternatively, old knowledge may have a larger role in learning, transfer, and consolidation due to being the primary model in the participants' repertoire (Székely et al. 2024).

If the encoding of new contingencies does not require the extinction or inhibition of old ones, then it would be computationally effective to reuse the (partial) knowledge that is already available (Heald et al. 2021; Székely et al. 2024). It has been shown that when participants unknowingly moved from a predictable

to an unpredictable environment, they continued using their old knowledge instead of abandoning it (Kóbor et al. 2020). Interestingly, when participants were asked to inhibit response sequences of the old knowledge, this even strengthened their expression when the old sequence returned. Therefore, not only can old and new memories of similar contingencies coexist (Szegedi-Hallgató et al. 2017; Horváth et al. 2022) but also new learning can even benefit from the existing imperfect memories (Székely et al. 2024).

Conclusion

In conclusion, we did not find evidence for competition between learning context (sequence) and knowledge (Ecker et al. 2007). The results align more closely with a predictive coding framework, in which sequence and stimulus properties are processed simultaneously (Apps and Tsakiris 2013; Blank et al. 2023). However, concomitant encoding is possible in a hierarchical organization, in which the higher-order context (sequence) level arbitrates the selection of the appropriate stimulus contingency model (Heald et al. 2021; Heald et al. 2023b; Székely et al. 2024). We propose that during rewiring, sequence-level encoding can be observed on the abstract coding level (C-cluster), and expression of the appropriate knowledge occurs through motor coding (R-cluster).

Author contributions

Ádám Takács (Data curation, Formal analysis, Methodology, Resources, Software, Visualization, Writing—original draft, Writing—review & editing), Teodóra Vékony (Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Writing—review & editing), Felipe Pedraza (Data curation, Investigation, Project administration, Writing—review & editing), Frederic Haesebaert (Investigation, Project administration, Writing—review & editing), Barbara Tillmann (Conceptualization, Writing—review & editing), Christian Beste (Methodology, Resources, Writing—review & editing), and Dezső Németh (Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Validation, Writing—review & editing).

Supplementary material

Supplementary material is available at *Cerebral Cortex* online.

Funding

This research was supported by the ANR Grant awarded within the framework of the Inserm CPJ (ANR-22-CPJ1-0042-01) (to D.N.); the IDEXLYON Fellowship of the University of Lyon as part of the Programme Investissements d'Avenir (ANR-16-IDEX-0005) (to D.N.); National Brain Research Program by the Hungarian Academy of Sciences (project NAP2022-I-1/2022) (to D.N.), and the Deutsche Forschungsgemeinschaft (DFG) TA1616/2-1 (to Á.T.).

Conflict of interest statement: None declared.

Data availability

We have made the code for computing the learning scores from the raw behavioral data, the reported behavioral data, EEG data, and analyses codes available on the Open Science Framework (https://osf.io/23hr8/?view_only=5bc367e3cbc049c39d8ef578546a5dd3). We can provide EEG data in various formats (raw, preprocessed, segmented, etc.) upon request. The study was not preregistered.

References

Apps MAJ, Tsakiris M. 2013. Predictive codes of familiarity and context during the perceptual learning of facial identities. *Nat Commun*. 4:1–10.

Aslin RN. 2017. Statistical learning: a powerful mechanism that operates by mere exposure. *WIREs Cogn Sci*. 8:e1373.

Bigdely-Shamlo N, Mullen T, Kothe C, Su KM, Robbins KA. 2015. The PREP pipeline: standardized preprocessing for large-scale EEG analysis. *Front Neuroinform*. 9:1–19. https://doi.org/10.3389/fninf.2015.00016.

Blank H, Alink A, Büchel C. 2023. Multivariate functional neuroimaging analyses reveal that strength-dependent face expectations are represented in higher-level face-identity areas. *Commun Biol*. 6:1–10. https://doi.org/10.1038/s42003-023-04508-8.

Delorme A, Makeig S. 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J Neurosci Methods*. 134:9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009.

Ecker UKH, Zimmer HD, Groh-Bordin C, Mecklinger A. 2007. Context effects on familiarity are familiarity effects of context - an electrophysiological study. *Int J Psychophysiol*. 64:146–156. https://doi.org/10.1016/j.ijpsycho.2007.01.005.

Éltető N, Nemeth D, Janacsek K, Dayan P. 2022. Tracking human skill learning with a hierarchical Bayesian sequence model. *PLoS Comput Biol*. 18:e1009866. https://doi.org/10.1371/journal.pcbi.1009866.

Farkas BC, Krajcsi A, Janacsek K, Nemeth D. 2024. The complexity of measuring reliability in learning tasks: an illustration using the alternating serial reaction time task. *Behav Res Methods*. 56:301–317. https://doi.org/10.3758/s13428-022-02038-5.

Fiser J, Aslin RN. 2001. Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychol Sci*. 12:499–504. https://doi.org/10.1111/1467-9280.00392.

Hallgató E, Gyori-Dani D, Pekár J, Janacsek K, Nemeth D. 2013. The differential consolidation of perceptual and motor learning in skill acquisition. *Cortex*. 49:1073–1081. https://doi.org/10.1016/j.cortex.2012.01.002.

Heald JB, Lengyel M, Wolpert DM. 2021. Contextual inference underlies the learning of sensorimotor repertoires. *Nature*. 600:489–493. https://doi.org/10.1038/s41586-021-04129-3.

Heald JB, Lengyel M, Wolpert DM. 2023a. Contextual inference in learning and memory. *Trends Cogn Sci*. 27:43–64. https://doi.org/10.1016/j.tics.2022.10.004.

Heald JB, Wolpert DM, Lengyel M. 2023b. The computational and neural bases of context-dependent learning. *Annu Rev Neurosci*. 46:233–258. https://doi.org/10.1146/annurev-neuro-092322-100402.

Horváth K, Nemeth D, Janacsek K. 2022. Inhibitory control hinders habit change. *Sci Rep*. 12:1–11.

Howard JH, Howard DV. 1997. Age differences in implicit learning of higher order dependencies in serial patterns. *Psychol Aging*. 12:634–656. https://doi.org/10.1037/0882-7974.12.4.634.

Kobor A, Janacsek K, Takacs A, Nemeth D. 2017. Statistical learning leads to persistent memory: evidence for one-year consolidation. *Sci Rep*. 7:1–10. https://doi.org/10.1038/s41598-017-00807-3.

Kóbor A et al. 2019. Tracking the implicit acquisition of nonadjacent transitional probabilities by ERPs. *Mem Cogn*. 47:1546–1566. https://doi.org/10.3758/s13421-019-00949-x.

Kóbor A, Horváth K, Kardos Z, Nemeth D, Janacsek K. 2020. Perceiving structure in unstructured stimuli: implicitly acquired prior knowledge impacts the processing of unpredictable transitional probabilities. *Cognition*. 205:104413. https://doi.org/10.1016/j.cognition.2020.104413.

Maheu M, Meyniel F, Dehaene S. 2022, 2022. Rational arbitration between statistics and rules in human sequence processing. *Nat Hum Behav*. 68:1087–1103. https://doi.org/10.1038/s41562-021-01259-6.

Mückschel M, Chmielewski W, Ziemssen T, Beste C. 2017. The norepinephrine system shows information-content specific properties during cognitive control – evidence from EEG and pupillary responses. *NeuroImage*. 149:44–52. https://doi.org/10.1016/j.neuroimage.2017.01.036.

Mullen T, Kothe C, Chi YM, Ojeda A, Kerth T, Makeig S, Cauwenberghs G, Jung TP. 2013. *Real-time modeling and 3D visualization of source dynamics and connectivity using wearable EEG*. In: Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, pp. 2184–2187. https://doi.org/10.1109/EMBC.2013.6609968, 2013.

Nemeth D, Janacsek K, Fiser J. 2013. Age-dependent and coordinated shift in performance between implicit and explicit skill learning. *Front Comput Neurosci*. 7:147. https://doi.org/10.3389/fncom.2013.00147.

Opitz A, Beste C, Stock AK. 2020. Using temporal EEG signal decomposition to identify specific neurophysiological correlates of distractor-response bindings proposed by the theory of event coding. *NeuroImage*. 209:116524. https://doi.org/10.1016/j.neuroimage.2020.116524.

Ouyang G, Zhou C. 2020. Characterizing the brain's dynamical response from scalp-level neural electrical signals: a review of methodology development. *Cogn Neurodyn*. 14:731–742. https://doi.org/10.1007/s11571-020-09631-4.

Ouyang G, Herzmann G, Zhou C, Sommer W. 2011. Residue iteration decomposition (RIDE): a new method to separate ERP components on the basis of latency variability in single trials. *Psychophysiology*. 48:1631–1647. https://doi.org/10.1111/j.1469-8986.2011.01269.x.

Ouyang G, Sommer W, Zhou C. 2015a. A toolbox for residue iteration decomposition (RIDE)-a method for the decomposition, reconstruction, and single trial analysis of event related potentials. *J Neurosci Methods*. 250:7–21. https://doi.org/10.1016/j.jneumeth.2014.10.009.

Ouyang G, Sommer W, Zhou C. 2015b. Updating and validating a new framework for restoring and analyzing latency-variable ERP components from single trials with residue iteration decomposition (RIDE). *Psychophysiology*. 52:839–856. https://doi.org/10.1111/psyp.12411.

Parra LC, Spence CD, Gerson AD, Sajda P. 2005. Recipes for the linear analysis of EEG. *NeuroImage*. 28:326–341. https://doi.org/10.1016/j.neuroimage.2005.05.032.

Pedroni A, Bahreini A, Langer N. 2019. Automagic: standardized preprocessing of big EEG data. *Neuro Image*. 200:460–473. https://doi.org/10.1016/j.neuroimage.2019.06.046.

Perrin F, Pernier J, Bertrand O, Echallier JF. 1989. Spherical splines for scalp potential and current density mapping. *Electroencephalogr Clin Neurophysiol*. 72:184–187. https://doi.org/10.1016/0013-4694(89)90180-6.

Petruo V, Takacs A, Mü M, Hommel B, Beste C. 2021. Multilevel decoding of task sets in neurophysiological data during cognitive flexibility. *ISCIENCE*. 24:103502. https://doi.org/10.1016/j.isci.2021.103502.

Pion-Tonachini L, Kreutz-Delgado K, Makeig S. 2019. The ICLabel dataset of electroencephalographic (EEG) independent component (IC) features. *Data Br*. 25:104101. https://doi.org/10.1016/j.dib.2019.104101.

Stark-Inbar A, Raza M, Taylor JA, Ivry RB. 2017. Individual differences in implicit motor learning: task specificity in sensorimotor adaptation and sequence learning. *J Neurophysiol*. 117:412–428. https://doi.org/10.1152/jn.01141.2015.

Szegedi-Hallgató E et al. 2017. Explicit instructions and consolidation promote rewiring of automatic behaviors in the human mind. *Sci Rep*. 7:4365. https://doi.org/10.1038/s41598-017-04500-3.

Szegedi-Hallgató E, Janacsek K, Nemeth D. 2019. Different levels of statistical learning - hidden potentials of sequence learning tasks. *PLoS One*. 14:e0221966. https://doi.org/10.1371/journal.pone.0221966.

Székely A, Török B, Nagy DG, Kiss MM, Janacsek K, Németh D, Orbán G. 2024. Identifying transfer learning in the reshaping of inductive biases. *Open Mind*. **8**:1107–1128.

Takacs A, Beste C. 2023. A neurophysiological perspective on the integration between incidental learning and cognitive control. *Commun Biol*. 6:1–13. https://doi.org/10.1038/s42003-023-04692-7.

Takács Á et al. 2021. Neurophysiological and functional neuroanatomical coding of statistical and deterministic rule information during sequence learning. *Hum Brain Mapp*. 42:3182–3201. https://doi.org/10.1002/hbm.25427.

Takács Á, Yu S, Mückschel M, Beste C. 2022. Protocol to decode representations from EEG data with intermixed signals using temporal signal decomposition and multivariate pattern-analysis. *STAR Protoc*. 3:101399. https://doi.org/10.1016/j.xpro.2022.101399.

Tóth-Fáber E, Janacsek K, Németh D. 2021. Statistical and sequence learning lead to persistent memory in children after a one-year offline period. *Sci Rep*. 11:1–11. https://doi.org/10.1038/s41598-021-90560-5.

Treder, M.S. 2020. MVPA-light: a classification and regression toolbox for multi-dimensional data. *Front Neurosci*. 14:289. https://doi.org/10.3389/fnins.2020.00289.

Tsivilis D, Otten LJ, Rugg MD. 2001. Context effects on the neural correlates of recognition memory: an electrophysiological study. *Neuron*. 31:497–505. https://doi.org/10.1016/S0896-6273(01)00376-2.

Turk-Browne NB, Scholl BJ, Chun MM, Johnson MK. 2009. Neural evidence of statistical learning: efficient detection of visual regularities without awareness. *J Cogn Neurosci*. 21:1934–1945. https://doi.org/10.1162/jocn.2009.21131.

Vékony T et al. 2023. Modality-specific and modality-independent neural representations work in concert in predictive processes during sequence learning. *Cereb Cortex*. 33:7783–7796. https://doi.org/10.1093/cercor/bhad079.

Widmann A, Schröger E, Maess B. 2015. Digital filter design for electrophysiological data - a practical approach. *J Neurosci Methods*. 250:34–46. https://doi.org/10.1016/j.jneumeth.2014.08.002.

Winkler I, Haufe S, Tangermann M. 2011. Automatic classification of Artifactual ICA-components for artifact removal in EEG signals. *Behav Brain Funct*. 7: 30–15. https://doi.org/10.1186/1744-9081-7-30.

Winkler I et al. 2014. Robust artifactual independent component classification for BCI practitioners. *J Neural Eng*. 11:035013. https://doi.org/10.1088/1741-2560/11/3/035013.

Wolff N, Mückschel M, Beste C. 2017. 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*. 222:3819–3831. https://doi.org/10.1007/s00429-017-1437-8.

Yonelinas AP. 1994. Receiver-operating characteristics in recognition memory: evidence for a dual-process model. *J Exp Psychol Learn Mem Cogn*. 20: 1341–1354. https://doi.org/10.1037/0278-7393.20.6.1341.

Cerebral Cortex, 2025, **35**, bhaf025
https://doi.org/10.1093/cercor/bhaf025
Original Article