Learning Amidst Noise: The Complementary Roles of Neural Predictive Activity and Representational Changes

Coumarane Tirou[1*], Teodóra Vékony[2], Oussama Abdoun[1], Laure Tosatto[3, 4], Andrea Brovelli[5], Marine Vernet[1], Dezső Németh[1, 2, 6†] & Romain Quentin[1†*]

[1]Lyon Neuroscience Research Center (CRNL), INSERM U1028, CNRS UMR5292, Université Claude Bernard Lyon 1, Lyon, France [2]Gran Canaria Cognitive Research Center, Atlántico Medio University, Las Palmas de Gran Canaria, Spain [3]CNRS, LPC, Aix Marseille Univ, Marseille, France [4]ILCB, Aix Marseille Univ, Aix-en-Provence, France [5]Institut de Neurosciences de la Timone UMR 7289, Aix Marseille Université, CNRS, 13005, Marseille, France [6]BML-NAP Research Group, ELTE Eötvös Loránd University & HUN-REN Research Centre for Natural Sciences, Budapest, Hungary

† Senior authors: these authors contributed equally to this work.
* Corresponding authors

Abstract

The ability to extract structured patterns from a noisy environment is fundamental to cognition, yet how the brain learns complex, non-adjacent regularities remains unclear. Using magnetoencephalography (MEG) during a visuomotor task, we tracked the neural dynamics as humans learned non-adjacent dependencies embedded in noise. We reveal that learning is supported by two temporally dissociable mechanisms. A rapid emergence of predictive activity, where neural patterns of expected stimuli appear before their onset, precedes measurable behavioral improvements. This is followed by a much slower representational change, characterized by an increased neural pattern similarity between linked, non-adjacent elements. Both processes are supported by a distributed consortium of sensorimotor, dorsal attention, salience, central executive, and cerebellar networks. These findings establish a temporal hierarchy for the neural mechanisms of learning, suggesting that fast predictions guide online behavior, which in turn facilitates the gradual consolidation of knowledge into stable neural representations.

Introduction

Learning allows humans to adapt to a constantly changing environment. Statistical learning plays a key role in this process as it enables the extraction of regularities from the environment without feedback or reward[1]. As a fundamental function of humans' cognition, such learning underlies a myriad of behaviors including language processing, skill acquisition, or perceptual decision making[2]. Statistical learning[3] has been extensively studied[3–6] and shown to be a domain-general mechanism operating across multiple modalities[7–11]. It plays a key role in natural language processing[3,11,12] as the hierarchical nature of language allows words to form relationships over varying distances within a sentence, resulting in non-adjacent dependencies[4,5]. Notably, such non-adjacent dependencies are considered crucial to distinguishing human language from other forms of communication systems and associated with humans' enhanced capacity to process higher hierarchical sequence structures compared to non-human animals[6]. Studies in humans have shown that extracting regularities is more challenging for non-adjacent dependencies than for adjacent dependencies[7,8], as sensitivity to non-adjacent dependencies develops later in infancy[7,9], and involves distinct neural networks with the lateral occipital, angular gyrus, medial and inferior frontal gyrus more involved in adjacent dependencies while the subcallosal cortex, paracingulate cortex, and anterior cingulate cortex were more involved in non-adjacent dependencies processing. Differences in the regions involved align with theories proposing that the learning of adjacent dependencies relies primarily on bottom-up signals from perceptual processing networks, whereas the learning of non-adjacent dependencies involves top-down signals from cognitive control networks to inhibit attention to intervening items and redirect it toward the relevant dependencies[10]. Recent advancements in neuroimaging and signal processing[7] have provided new insights into the neural mechanisms involved in statistical learning of adjacent dependencies, namely representational change and predictive activity[12–15]. However, it remains unclear whether these mechanisms support statistical learning of non-adjacent dependencies[11], and how they dynamically emerge during learning.

Neural representational change is defined as the increase in similarity of neural activity patterns between associated stimuli[12,13,15,16]. A seminal study in monkeys[17] demonstrated that neurons in the inferior temporal lobe developed similar responses to images presented in close temporal proximity, suggesting that neural selectivity can emerge from learned temporal associations and may reflect a mechanism for statistical learning. A similar effect has been observed in humans using fMRI[13], showing that passive exposure to structured and random sequences of abstract images led to increased neural pattern similarity for structured versus random image pairs in the medial temporal lobe and hippocampus. Studies examining changes in neural similarity typically focused on post-learning effects, without measuring behavioral evidence of learning (e.g., reaction time or error rates). They did not track neural changes as learning unfolded, nor did they link these changes to behavioral markers of learning[18].

Neural predictive activity or predictive coding[19,20] theories posit that brains are known to be predictive machines[19] that constantly try to match incoming sensory inputs. The Bayesian brain theory of predictive coding postulates that the brain constructs an internal model of the world by detecting statistical regularities from sensory inputs[21]. Under this theory, the brain has priors on the environment based on its internal model that it uses to make predictions (i.e., top-down modulations) about upcoming stimuli. When an unexpected event arises, it makes use of prediction errors to update its internal model (i.e., bottom-up modulations) thereby continuously improving its predictions. Predictive coding has been extensively studied across sensory modalities with neural recordings[22,23]: distinct brain responses to a deviant stimulus in (e.g., visual or auditory) sensory regions. Some studies[24,25] suggest that the brain pre-activates neural ensembles involved in the perception of an upcoming, predictable stimulus before it appears (i.e., predicting a tone at 440 Hz activates similar patterns of brain activity than perceiving it). However, evidence for this hypothesis remains scarce, and methodological limitations due to high physical proximity between stimuli may have hindered several studies[24,25]. Similar to neural representational changes, it remains unclear whether predictive activity (perception-like or not) emerges during the learning of non-adjacent dependencies and how it builds up over time during learning.

Using an explicit probabilistic learning task with relevant information embedded in noise, combined with magnetoencephalography (MEG) recordings, we identified both representational changes and predictive neural activity throughout learning. Representational changes were primarily observed between 300 and 550 ms after stimulus onset, within the sensorimotor, dorsal attention, central executive, and cerebellar networks. Predictive activity preceding the onset of the stimulus were observed in the same brain regions, except for the central executive network. Neural representational changes correlated with learning and predictive activity emerged before the neural representational changes. Our results suggest that both neural representational changes and predictive activity support the statistical learning of non-adjacent dependencies, and reveal the dynamics of these mechanisms as learning unfolds.

Results

We recorded MEG from 15 human participants as they performed the cued Alternative Serial Reaction Time (cASRT) task, responding to a sequence of arrows by pressing the corresponding buttons. Task blocks included repeating (yellow) and random (red) arrows (Fig. 1a), allowing the emergence of predictable (i.e., *pattern*) and unpredictable (i.e., *random*) elements (Fig. 1b). Participants first completed three practice blocks in which no sequence was embedded in the yellow arrows, followed by four sessions of five learning blocks each. They were informed that the yellow stimuli followed a fixed sequence in learning blocks and were instructed to identify it and optimize their performance using this knowledge.

Average reaction time (RT) per session for all trials decreased over time (Fig. 1c) (main effect of session: $F_{3, 36} = 15.22$, $p < 0.001$), demonstrating that participants are getting faster independently of trial conditions. Average RT was smaller for pattern pairs than random pairs (main effect of trial condition: $F_{1, 14} = 21.35$, $p < 0.001$), demonstrating that participants became faster for pattern pairs compared to random pairs. Finally, the difference in RT between pattern and random pairs (i.e., learning index) increased over session (interaction of session and trial condition: $F_{3, 36} = 6.87$, $p < 0.01$). These results confirm that statistical learning has occurred during the task. The learning index was calculated for each session, excluding the practice session (set to 0). The learning index was significantly above 0 for session 1 ($p < 0.05$) and sessions 2 to 4 (all $p < 0.01$).

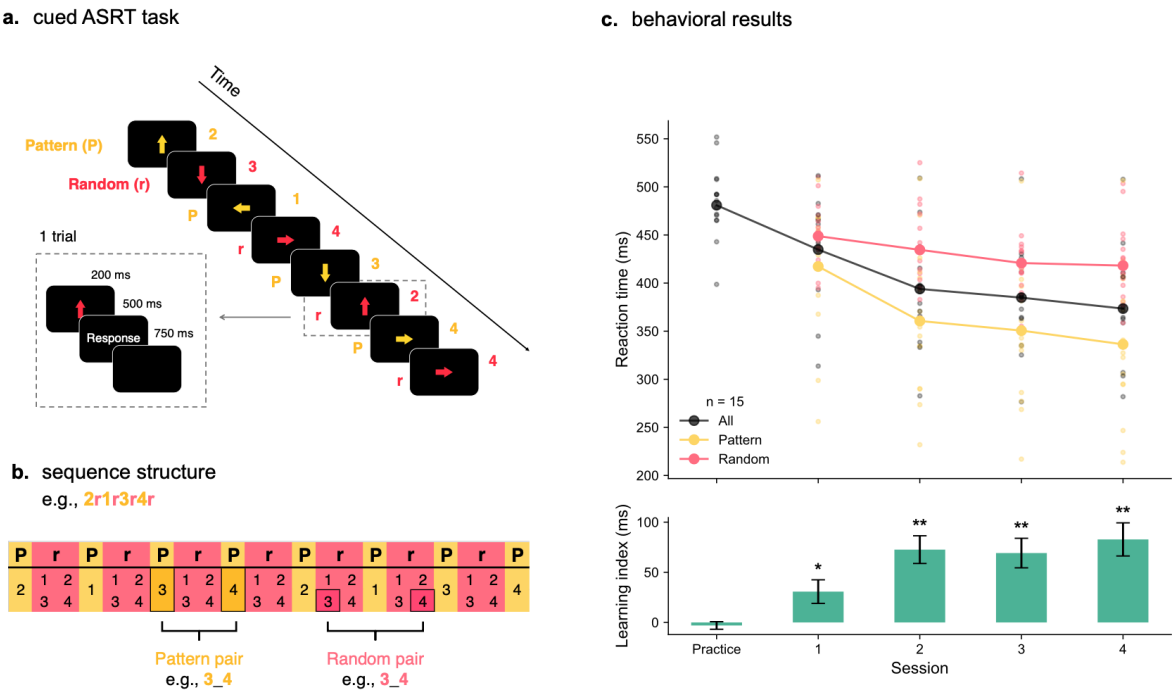

Fig. 1. The cued Alternating Serial Reaction Time (cASRT) task leads to significant statistical learning of non-adjacent dependencies. a In the cASRT task, an arrow stimulus appears in the center of the screen, alternating between repeating pattern trials (yellow) and unpredictable random trials (red). Each trial displays the arrow for 200 ms, followed by a 500 ms response window for participants to press the corresponding button. The inter-trial interval is 750 ms. **b** The alternating sequence structure uses numbers to denote spatial directions (1 = left, 2 = up, 3 = down, 4 = right). Statistical learning is measured as the reaction time (RT) in ms difference between pattern and random pairs. **c** Participants' (n=15) RTs for pattern stimuli and random stimuli in the cASRT task. Scatter points represent RTs per stimulus type for each participant, while tick points and lines show the average RTs across participants per session. Each session comprises five blocks of 85 trials each of cASRT, except for the practice session, which includes three blocks with no sequence in the pattern trials. The bottom panel illustrates the learning index (ms) per session which is the difference in RT between random and pattern pairs. Error bars indicate the SEM. Asterisks denote statistical significance (**: $p < 0.01$, one-sample *t*-test). Participants were significantly faster for the pattern than for random stimuli, and this difference increased across sessions, confirming statistical learning.

Late representational change after stimulus onset at the brain level correlates with behavioral performance.

We conducted representational similarity analysis (RSA) using a Leave-one-Block-out (LoBo) cross-validated Mahalanobis distance (cvMD)[26,27] between elements of both pattern and random pairs (Fig. 2a). Then, the distance within pattern pairs was subtracted from that of the random (see formula in *Methods* section). This yielded a similarity index that reflects changes in similarity within pattern pairs relative to changes in similarity within random pairs (Fig. 2b). The time course of this similarity index revealed a significant positive cluster ranging from 0.38 to 0.43 s post-stimulus onset (cluster level, $p < 0.05$ corrected). To investigate the link between learning and such representational change, we calculated, for each participant, the correlation coefficient between the similarity index and the learning index across learning blocks (Fig. 2c). The Spearman rhos across participants revealed two significant above 0 clusters from 0.31 to 0.36 s and 0.47 to 0.50 s post-stimulus onset (cluster level, $p < 0.05$ corrected).

To estimate the average effect, we computed the representational change by averaging the similarity index values within a 0.30-0.55 s time window for each participant. The average of similarity index yielded a significant above 0 effect (one-sample *t*-test, $p = 0.04$). The same was done for the Spearman rho's values. This mean representational change was also positively correlated with learning index ($p = 0.02$, one-sample *t*-test on the Spearman rhos against zero) (Fig. 2d).

Overall, these findings demonstrate that the brain updates its representations of pattern pairs relative to random pairs by increasing the similarity of non-adjacent elements within pattern pairs, while also disregarding the non-informative random elements between them. The positive correlation with the amount of learning suggests that such representational change could drive the learning process.

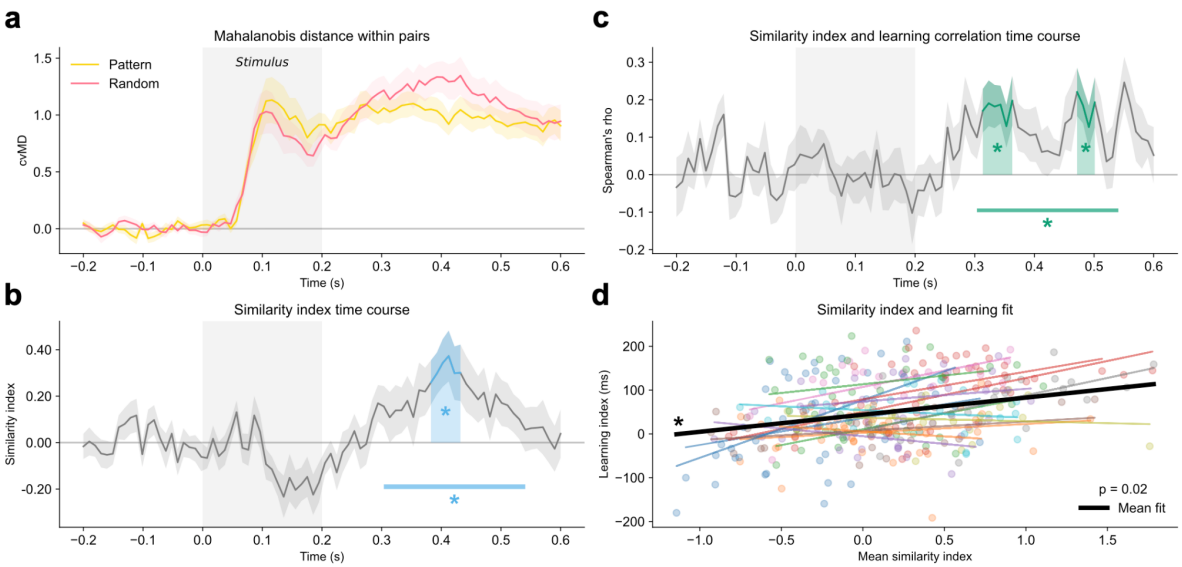

Fig. 2. Representational similarity analysis (RSA): late representational change positively correlates with learning. a Time course of the Leave-one-Block-out cross-validated Mahalanobis distance (cvMD) for pattern and random pairs, averaged across participants. The grey rectangle area indicates the stimulus duration (200 ms). **b** Time course of the similarity index, averaged across participants and blocks. Solid lines and shaded areas indicate the mean and the SEM, respectively. Asterisk and color-filled areas indicate statistical significance (cluster-level, $p < 0.05$ corrected). Elements within pattern pairs evoked brain activities more similar to each other relative to elements within random pairs from 0.30 to 0.55 s post-stimulus onset (colored horizontal line) with a significant cluster at 0.38-0.43 s. **c** Time course of Spearman's rho correlation between the similarity index and the learning index across sessions. Solid lines and shaded areas indicate the mean and the SEM, respectively. Asterisk and color-filled areas indicate statistical significance (cluster-level, $p < 0.05$ corrected). The similarity index positively correlates with learning during 0.30 to 0.55 s post-stimulus onset (colored horizontal line) with significant clusters at 0.31-0.36 and 0.47-0.50 s. **d** Relationship between the mean representational change and the learning index across blocks. The mean representational change was calculated by averaging the similarity index between 0.30 and 0.55 s for each session and each participant. Each participant's fit is depicted in different colors with solid color lines indicating the mean fit across blocks. The bold black line shows the mean fit across participants and sessions. Within-subject computed rhos are significantly different from 0 ($p = 0.02$, one-sample *t*-test), confirming that increased learning is associated with greater representational change.

Distributed representational change across the sensorimotor, dorsal attention, central executive, and cerebellar networks.

To identify the brain spatial localization of the representational change effect, we performed source reconstruction of our MEG data using beamforming[28,29]. The participants' brains were parcellated into 200 cortical regions, which were further grouped into seven networks: visual, sensorimotor, dorsal attention, salience, limbic, central executive, and default mode networks. Additionally, we incorporated three subcortical structures: the hippocampus, thalamus, and cerebellum. These regions served as our regions of interest (ROIs). RSA was performed in each network using the same LoBo cvMD distance measure as in the previous sensor-space analysis. While only the dorsal attention and central executive networks showed significant continuous clusters post-stimulus onset (cluster level, $p < 0.05$ corrected), averaging the similarity index in a time window ranging from 0.30 to 0.55 s post-stimulus onset (i.e., corresponding to the timing of the mean effect at the sensor level) revealed a significant mean effect in the the sensorimotor, dorsal attention, central executive, and cerebellar networks (one-sample *t*-test, $p < 0.05$) (Fig.3b). The time course correlation between the similarity index with learning revealed no significant clusters across any ROIs (cluster level, all $p > 0.05$ corrected), but the sensorimotor network averaged during our time window of interest revealed a significant above 0 effect (one sample *t*-test, $p < 0.05$) (Fig. 3c).

In summary, these findings indicate that the initial representational change observed at the brain level is distributed across the sensorimotor, dorsal attention, central executive, and cerebellar networks at the source level. No evident correlation was found between the magnitude of learning and representational change across these networks, except for the sensorimotor network.

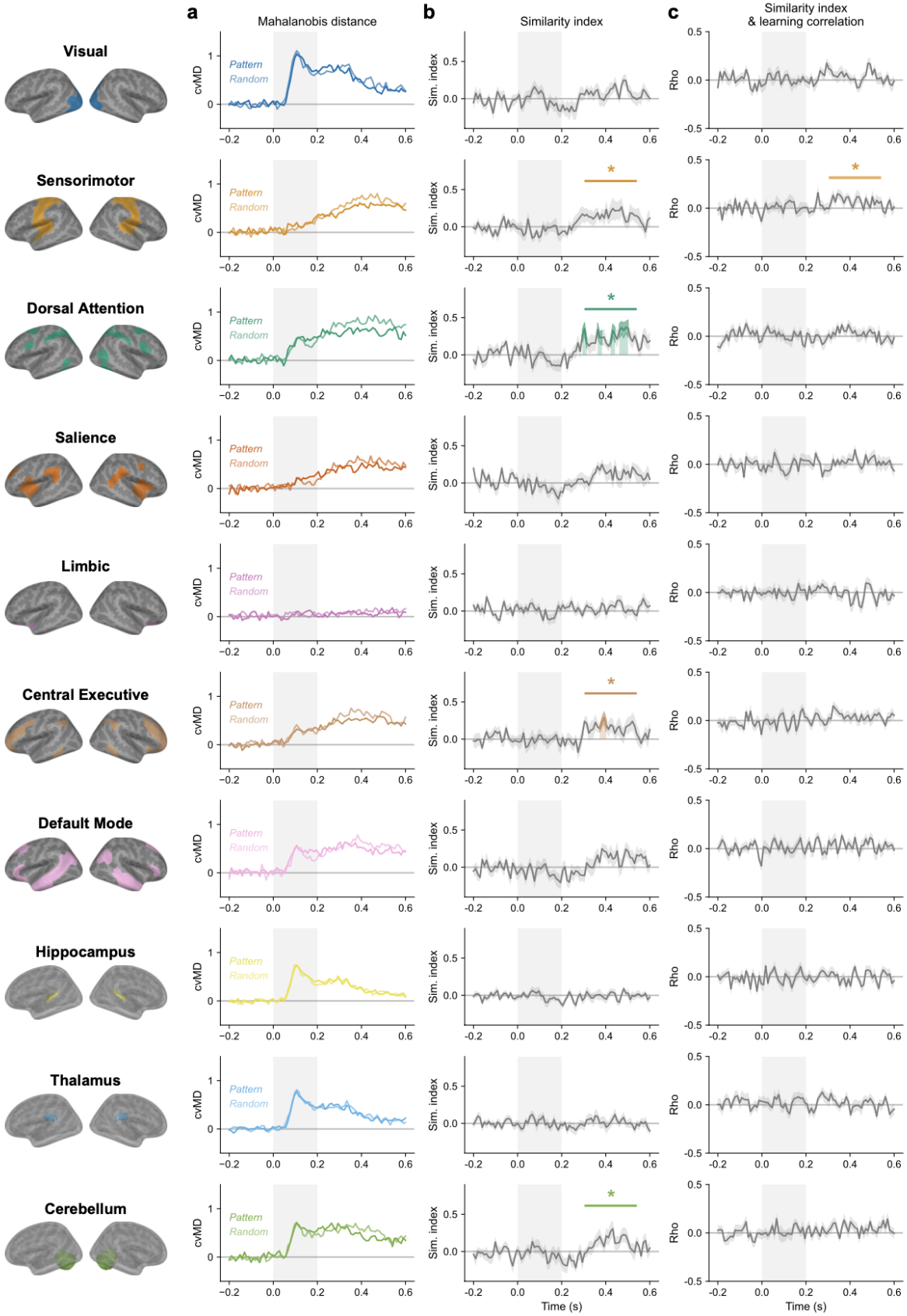

Fig. 3. Source space RSA: distributed representational change in the sensorimotor, dorsal attention, central executive, and cerebellar networks. a Leave-one-Block-out cross-validated Mahalanobis distance for pattern and random elements. Distances were averaged across participants. The grey rectangle area indicates the duration of the stimulus (200 ms). Color-

filled areas indicate statistical significance (cluster-level, $p < 0.05$ corrected). **b** Similarity index time course across ROIs, averaged across sessions and participants. Color-filled areas indicate statistical significance (cluster-level, $p < 0.05$ corrected). Colored horizontal lines above curves indicate above 0 statistical significance for values averaged across 0.30-0.55 s time window (one sample *t*-test, $p < 0.05$). Significant clusters of representational change were found in the dorsal attention and central executive networks. The average of similarity index revealed significant mean representational change effect in the sensorimotor,, dorsal attention, central executive, and cerebellar networks (one sample *t*-test, $p < 0.05$). **c** Spearman correlation time course of the similarity index and learning index across sessions. Colored lines above curves indicate above 0 statistical significance for values averaged across 0.30-0.55 s time window (one sample *t*-test, $p < 0.05$). No significant clusters were found in any of the ROIs (cluster-level, all $p > 0.05$ corrected). The sensorimotor network exhibited a significant above 0 correlation effect.

Information about the upcoming stimulus is present before its appearance.

To investigate predictive activity, we used a Logistic Regression classifier with LoBo cross-validation strategy on both random and pattern trials and the time generalization[30] method to capture mental representations and investigate if they generalize before the stimulus appearance. This analysis showed stable representations for both trial types, with sustained representations before the stimulus onset for pattern trials but not for random trials (cluster level, all $p < 0.05$ corrected) (Fig. 3a). When contrasted (i.e., Time generalization matrix of pattern trials – matrix of random trials), we found a significant cluster of the neural representation of the upcoming stimulus early ~500 ms before its appearance, corresponding to the inter-trial interval, and building up to the stimulus onset (Fig. 3b, top, black rectangle). When correlated with the learning performance, the contrast revealed no significant cluster (cluster level, all $p > 0.05$ corrected) (Fig. 3b, bottom).

In summary, we found a significant predictive activity of the upcoming stimulus around 500 ms before its appearance during highly predictive trials.

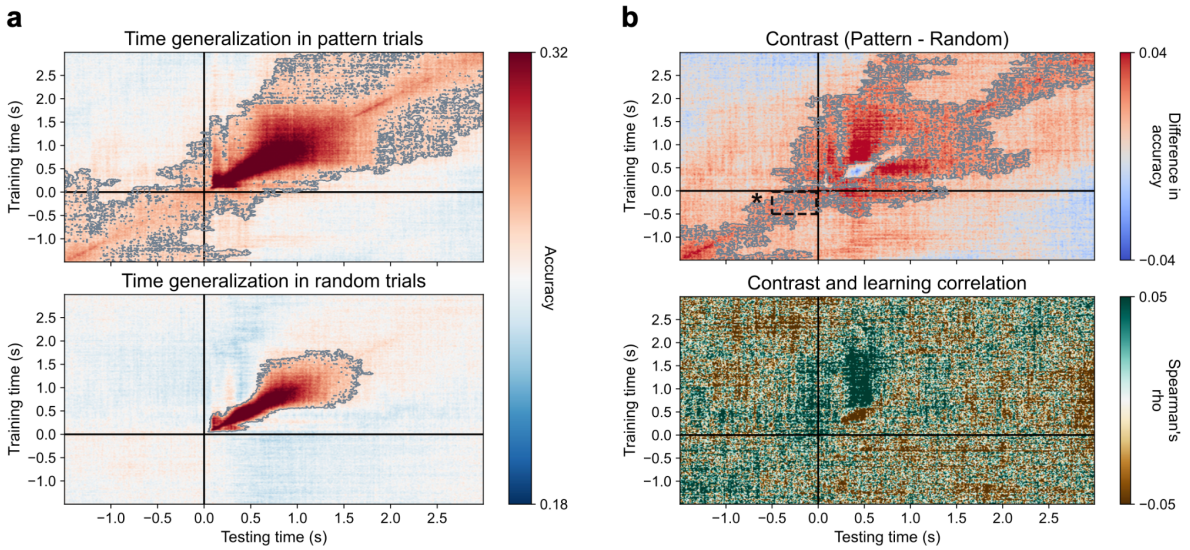

Fig. 4. Temporal generalization: a predictive activity effect of expected stimuli in pattern trials. a Time generalization of spatial arrows during pattern and random trials. Black lines indicate the stimulus onset timing (0 s). Statistically significant temporal clusters ($p < 0.01$) are delineated by a grey line. We found sustained neural representations in the post-stimulus period for both pattern and random trials, but also in the pre-stimulus period in pattern trials. **b** The contrast between pattern and random trials time generalization matrices (top) returned a significant cluster in the inter-trial interval before the stimulus onset (black square) demonstrating predictive activity in pattern trials. The Spearman correlation between the contrast matrix and learning index returned no significant clusters.

Distributed predictive activity in the sensorimotor, dorsal attention, salience, and cerebellar networks.

To investigate predictive activity at the source level, we employed the same source reconstruction pipeline as for the RSA. Sensor-space analyses revealed that the effect was strongest and statistically significant mainly along the diagonal preceding stimulus onset: reflecting training and testing at identical pre-stimulus time points (i.e., decoding). Therefore, only these segments were considered for further analyses. Decoding across blocks revealed sustained neural representations (cluster level, all $p < 0.05$

corrected) during pre- and post-stimulus onset periods in the sensorimotor, dorsal attention, and salience networks (Fig. 4a). In random trials, neural representations across sessions were only found after stimulus onset in all ROIs (Fig. 5a, middle) (cluster level, all $p < 0.05$ corrected). We then computed the contrast (i.e., the difference between pattern and random trials decoding performance), which displayed significant clusters in the pre-stimulus period in the sensorimotor and dorsal attention networks only (Fig. 4a, right) (cluster level, all $p < 0.05$ corrected). We then computed the average of the contrast values within a 500 ms time window preceding stimulus appearance, which revealed significant effects in the sensorimotor, dorsal attention, salience, and cerebellar networks (one sample *t*-test, $p < 0.05$).

In summary, our findings show a distributed predictive activity building up to the stimulus onset across the sensorimotor, dorsal attention, salience, and cerebellar networks.

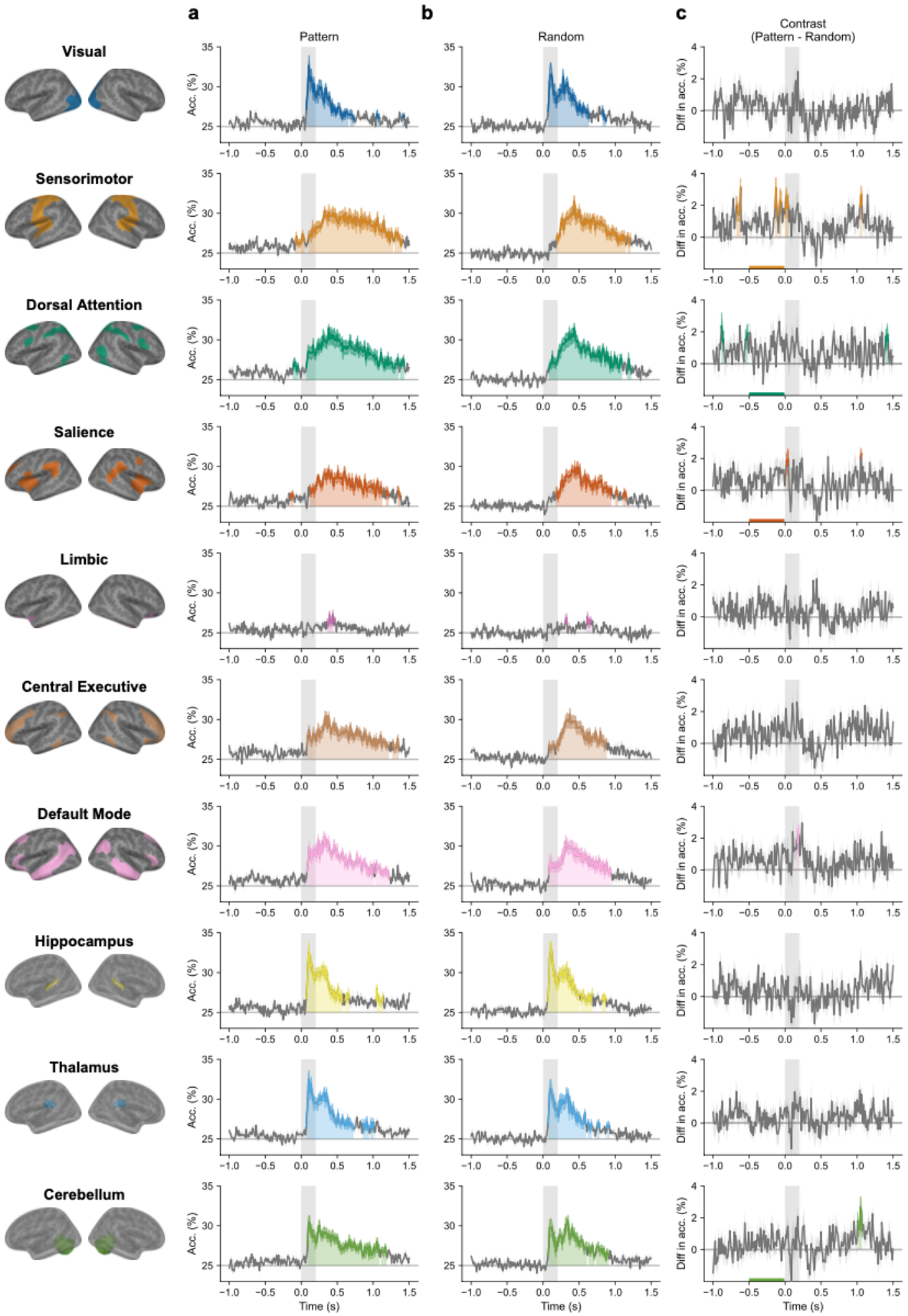

Fig. 5. Source space decoding: a distributed predictive activity across sensorimotor, dorsal attention, salience, and cerebellar networks. a Decoding performance of spatial arrows in pattern trials across all ROIs, using a Logistic Regression classifier and a Leave-one-Block-out cross-validation, averaged across participants. The grey rectangle area indicates the

stimulus duration (200 ms). Statistically significant temporal clusters ($p < 0.05$) are delineated by colored lines. Sustained neural representations were found during pre-stimulus period in pattern trials in all ROIs, both the pre- and post-stimulus period in pattern trials in the sensorimotor and dorsal attention networks, and during the post-stimulus period only in random trials. **b** Same as panel *a* but with random trials. Sustained neural representations were found during the post-stimulus period only in random trials across all ROIs. **c** The contrast between pattern and random trials decoding performance returned significant clusters in the sensorimotor and dorsal attention networks only. Average values within 500 ms preceding stimulus revealed significant effects in sensorimotor, dorsal attention, salience, and cerebellar networks (colored horizontal lines, one sample *t*-test, $p < 0.05$).

Non-linear temporal modeling on behavior, predictive activity, and representational change reveals distinct temporal dynamics across blocks and ROIs.

In order to track changes in representations and predictions over the course of learning, we applied non-linear temporal modeling to the block-wise mean effects of representational changes, predictive activity, and behavioral performance (i.e., learning index) (Fig. 6). Behavioral performance increased from block 6 to block 10 before reaching a plateau and was significantly above 0 from block 8 onward ($p < 0.0001$). Predictive activity rose from the first block itself until block 6 before reaching a plateau and was significantly above 0 from block 4 onward ($p < 0.05$). Representational change increases throughout the behavioral task (block 1 to 23) and is significantly above 0 from block 11 until the end of experiment ($p < 0.05$). In summary, our results show that behavioral performance, predictive activity, and representational change have different temporal dynamics as learning unfolded.

We conducted the same analysis with mean effects of representational change and predictive activity in source space across all our ROIs (Fig. 7). For predictive activity, we found significant above 0 effects in the sensorimotor and dorsal attention networks from block 1 onward, and salience network from block 1 to block 19 (all $p < 0.001$). When uncorrected for multiple comparisons, we found significant effects in the central executive from block 8 to block 21, the default mode network from block 11 to block 21, and the cerebellum from block 12 to 23 (all $p < 0.05$). For representational change, we found significant above 0 effects in the sensorimotor network from block 12 onward ($p < 0.01$) and the dorsal attention network from block 8 onward ($p < 0.0001$). When uncorrected for multiple comparisons, we found significant effects in the salience network from block 12 onward, the central executive network from block 10 to block 22, and the cerebellum from block 12 onward (all $p < 0.05$). In summary, our results show that representational change and predictive activity have different temporal dynamics mainly distributed across the sensorimotor and attentional networks.

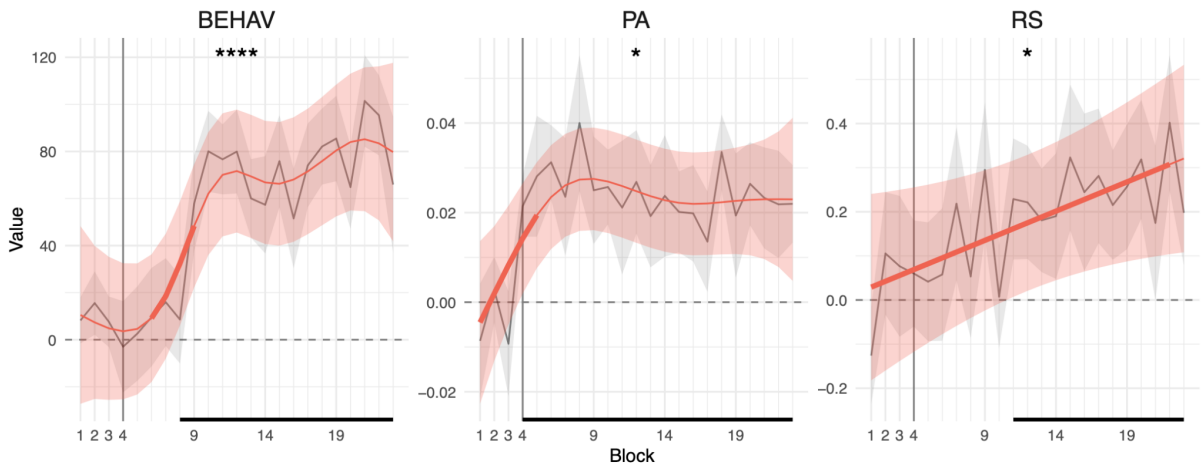

Fig. 6. Behavior, predictive activity and representational change have different temporal dynamics across blocks. Jagged (grey) lines and ribbons are group mean +/- SEM calculated from data. Smooth (red) lines and ribbons are the model's estimated marginal means and 95% CI. Asterisks indicate significance (H0 = 'temporal trend is constant and equal to 0'). Vertical grey lines indicate the end of practice blocks and beginning of learning blocks. Horizontal segments indicate blocks significantly different from 0. Thicker portions of the model smooth indicate significant temporal slopes (derivative). BEHAV: behavior, PA: predictive activity, RS: representational change. ****: p < .0001, *: p < .05.

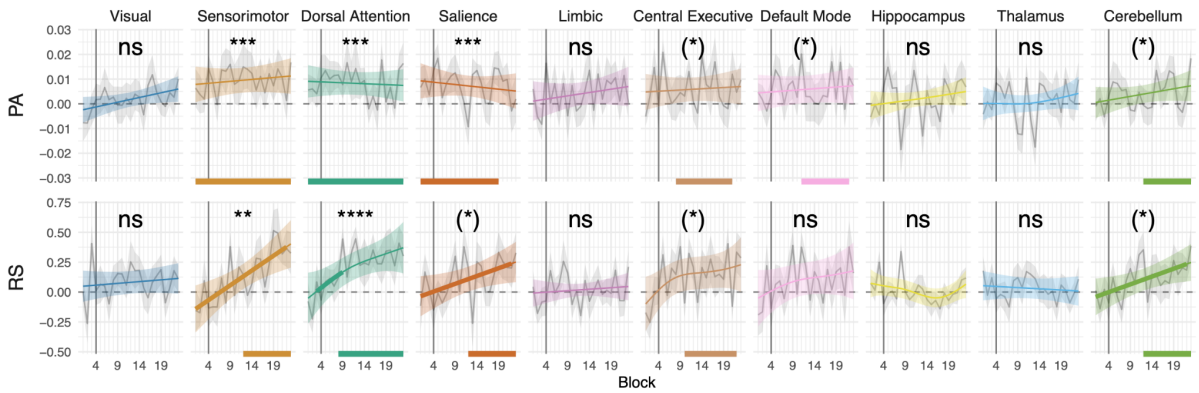

Fig. 7. Distributed predictive activity and representational change effects across blocks in sensorimotor and attentional networks. Jagged (grey) lines and ribbons are group mean +/- SEM calculated from data. Smooth colored lines and ribbons are models' estimated marginal means and 95% CI. Asterisks indicate significance (H0 = 'temporal trend is constant and equal to 0'). Vertical grey lines indicate the end of practice blocks and beginning of learning blocks. When statistical significance does not survive a per-metric Holm correction for multiple comparison, it is displayed between parentheses. Horizontal, colored segments indicate blocks significantly different from 0. Thicker portions of the model smooth indicate significant temporal slope (derivative). BEHAV: behavior, PA: predictive activity, RS: representational change. ****: $p < .0001$, ***: $p < .001$, **: $p < .01$, *: $p < .05$, ns: non significant.

Discussion

To adapt to a dynamic world, the human brain must extract stable regularities from noisy environments, a process known as statistical learning[1,3]. While the neural basis of learning simple adjacent dependencies has been explored[13–15], how the brain learns complex, non-adjacent patterns separated by irrelevant information remains poorly understood. Here, using MEG during a visuomotor task, we demonstrated that learning non-adjacent dependencies is supported by two neurally and temporally dissociable mechanisms: predictive activity and representational change. We first reveal a rapid emergence of predictive activity, where neural patterns for expected stimuli appear up to 500 ms prior to their onset across a distributed network including sensorimotor, dorsal attention, salience, central executive, and cerebellar regions. Second, we found a more gradual representational change, where the neural patterns evoked by linked, non-adjacent stimuli become more similar. This representational shift occurs much later, between 300-550 ms after stimulus onset, and is localized to a similar set of sensorimotor, attention, executive, and cerebellar networks. Critically, our temporal modeling reveals a clear hierarchy: predictive activity arises first, followed by behavioral improvements, which in turn precede the consolidation of representational change that continues to increase throughout the task. These results uncover the distinct temporal dynamics of the core neural mechanisms supporting learning, suggesting that fast predictions guide initial behavior while slower representational plasticity solidifies learned structures amidst noise.

Schapiro et al.[13,15] evidenced representational change as a mechanism of statistical learning of adjacent dependencies using a passive exposure task. Their fMRI study revealed the hippocampus and the medial temporal lobe as areas that support this mechanism. Our results extend the findings of Schapiro et al.[13] by showing that representational change also supports statistical learning of non-adjacent dependencies, indicating that this mechanism remains robust when regularities are embedded with noisy information. By incorporating a motor response and leveraging the high-temporal resolution of MEG, our study provided a more detailed account of statistical learning and revealed a positive link between representational change and behavioral performance. However, our findings differ from Schapiro et al.[13] in three key ways. First, we did not find evidence of representational change in the hippocampus. This result could stem from the inherent challenges of resolving subcortical signals in MEG source space, where signal-to-noise ratios are typically low. Second, Schapiro et al.[13] reported that representational change emerged after an hour of stimuli exposure[15] while our study shows that the effect is present after about 17 mins (from block 11) after the sequence is introduced. The rapid

emergence of the effect may be due to the inclusion of a motor response, which likely increased participants' engagement and active involvement in the task compared to passive exposure, further highlighting the link between behavioral performance and representational change that we found. Additionally, participants were aware of the presence of a hidden sequence within the stimulus stream, which may have facilitated learning. Third, we found representational change of stimuli that are visually similar whereas Schapiro *et al.* did not find any between visually similar stimuli. This difference could be explained by the motor component of our task that enhances engagement. Interestingly, our representational change effect was found during a time frame that corresponds to the RT range. This aligns with the effect observed in the sensorimotor network and its correlation with behavioral performance, suggesting that the motor representations of associated stimuli became more similar while visual representations remained unaffected, likely requiring greater dissimilarity between stimuli. We also build upon previous temporally resolved RSA studies[12,14], that identified representational changes for predictable auditory stimuli, extending these findings to visual perception and the learning of non-adjacent dependencies. Our results contrast with a late effect (i.e., 300 to 550 ms post-stimulus onset) compared to an early effect (i.e., around 120 ms post-stimulus onset) of previous studies, which could be attributed to the experimental paradigm that incorporates a motor component. Further to this, our results of representational change show a distributed mechanism across various networks rather than a localized one. This is in line with previous works[12] suggesting that statistical learning engages distributed cortical networks involved in the parallel encoding of both low (e.g., transitional probabilities) and high level features (e.g., sequence order and identity).

Extensive evidence supports that the brain continuously seeks to match incoming perceptual input through predictive activity shaped by previous experience. Despite neural mechanisms underlying such predictive activity being unclear, one widely accepted hypothesis is that predictive signals are transmitted through top-down pathways from higher-order to sensory regions[20,25]. As such, pre-activation of neural ensembles that encode a stimulus before its appearance, has been a potential candidate for neural prediction. Previous work in humans investigating this mechanism[24,25] have since been challenged on methodological grounds, as the high physical similarity between stimuli can produce spurious effects that are difficult to distinguish from true prediction. The structure of the cASRT task - where each preceding (*n-1*) trial is always uninformative of the current (*n*) trial - provides a unique opportunity to isolate anticipatory neural activity that can be more confidently interpreted as predictive activity. Using the temporal generalization method[30] in sensor space, we sought to identify predictive activity. First, when the classifier was both trained and tested within the pre-stimulus period, we observed a significant cluster emerging around 500 ms before and building up to stimulus appearance, again only in pattern trials. This suggests top-down modulation of sensory activity from higher-order regions in anticipation of predictable stimuli, but could also be a pre-activation of neural ensembles that code for stimulus perception. Second, our results show a continuous decoding from the *n-2* stimulus to the current stimulus *n* during pattern trials, indicating the presence of a sustained memory trace when two stimuli are temporally linked. Notably, such a memory trace is maintained despite an uninformative interference (i.e., stimulus *n-1* is a random trial), further supporting the robustness of sequential memory representations. This finding additionally suggests that neural ensembles are selectively tuned to relevant (pattern) information while effectively filtering out irrelevant (random) noise. Next, we sought to explore the spatial localization of these neural predictive processes among 7 networks and additional 3 subcortical regions. Out of all our ROIs, the sensorimotor, dorsal attention, salience, and cerebellar networks revealed a predictive activity within the same 500 ms pre-stimulus period. These results show a distributed predictive modulation activity across not only sensory networks (i.e., sensorimotor) but also higher-order networks (e.g., dorsal attention, salience, and cerebellum) indicating that top-down modulation is not limited to sensory regions, but may also involve dynamic interactions among higher-order cognitive networks. Future research could explore the functional connectivity between these regions, which, although highly relevant, falls outside the scope of the current work.

Quantifying effects of behavioral performance, predictive activity, and representational change across blocks revealed distinct dynamics for each process. Predictive activity emerged first, becoming

significant from the first block of learning (block 4), followed by the learning index, and finally representational change. No direct correlation was found between predictive activity and learning index (Fig. 4, *bottom panel*), despite predictive activity preceding the learning index as revealed by GAMM analysis (Fig. 6). This suggests a non-linear relationship between predictive activity and behavioral performance, which additive non-linear modeling captures. One hypothesis would be that predictive activity is a necessary but not a sufficient condition for learning. The sequence of predictive activity preceding the learning index, followed by representational change, may reflect a 'fast logic' phenomenon, consistent with dual-process theory[31] which posits the coexistence of intuitive and deliberate modes of thinking. Although participants were aware of the presence of a hidden sequence in the task, the brain may have detected these regularities implicitly, with explicit sequence knowledge emerging later and subsequently translating into improved behavioral performance and increased representational similarity of pattern pairs. Another hypothesis is that predictive activity may drive representational change, which in turn supports behavioral improvement. Early, rapid predictive activity in sensorimotor and attentional networks could facilitate initial performance gains. This sustained prediction then provides a stable signal to gradually reshape neural representations. Slower representational change, particularly in attentional networks, may consolidate and further optimize learned behavior, explaining why representational change continues to rise even after behavioral performance plateaus. At the source level, representational change in the dorsal attention network became significant from block 8 onward, coinciding with the rise of the learning index. Predictive activity in the same network appeared from block 1 onward, suggesting, pending confirmation in future studies, a potential causal effect of predictive activity on representational change during statistical learning. Together, these findings support the engagement of large-scale neural interactions in predictive processing and representational change in response to environmental statistical regularities[14]. Future studies could investigate the causal relationship between predictive activity and representational change to better understand their interplay.

The role of attentional networks has been strongly linked to non-adjacent dependency processing involving top-down signals from cognitive control networks to inhibit attention to intervening items and redirect it toward the relevant dependencies[10]. Our results in representational change and predictive activity corroborate these previous studies. Indeed, dorsal attention and salience networks are involved in predictive activity while representations in the dorsal attention network exhibit representational change (Fig. 3 & 5). The dorsal attentional network mediates top-down internal guidance of information based on prior knowledge, willful goals and plans whereas the salience (or ventral attention) network mediates bottom-up signals assumed to be involved in detecting unattended or unexpected stimuli and triggering shifts of attention[32,33]. It is interesting to notice that the dorsal attention network is involved throughout all blocks compared to the salience network which is involved until block 19 (Fig. 7). These results hint at a cooperation of both attentional systems to process stimuli[33,34] with a disengagement of the salience networks toward the end of the task as the sequence structure is completely assimilated by the participants whereas the dorsal attention is active throughout manifesting its involvement in prediction based on prior knowledge.

Our findings also resonate with recent large-scale and electrophysiological accounts of statistical learning. Functional connectivity work has shown that statistical learning is facilitated by a reduction in frontal connectivity, particularly between the superior frontal gyrus and attentional or salience networks, suggesting that diminished top-down control may enable the extraction of novel regularities from noisy input[35]. Complementing this, electrophysiological evidence indicates that perceptual and motor codes contribute jointly to the predictive processes underlying sequence learning, supporting the view that probabilistic learning relies on the concerted operation of both modality-specific and modality-independent neural representations[36]. Taken together, these results converge with our MEG data by highlighting that predictive activity and representational change emerge not in isolation but through the dynamic coordination of distributed brain systems that flexibly balance top-down control and multi-level coding to optimize learning.

While this study provides novel insights into the distinct temporal dynamics and spatial localization of representational change and predictive activity, it also has certain limitations. First, the relatively small sample size may have amplified or attenuated neural effects, particularly in source space. However, the use of powerful within-subject statistical methods (e.g., non-parametric cluster-based permutation tests) and advanced modeling (GAMMs) provides robust detection of effects at the individual level. Second, the use of arrows as stimuli introduces potential confounds: participants may hold pre-existing representations of arrows, and their physical similarity—differing only in pointing direction—could bias true representational change effects, at least in its temporal dynamics. Third, the explicit nature of the task, in which participants were instructed to identify and exploit the hidden sequence to optimize performance, may have artificially enhanced the temporal dynamics of the observed effects, especially for predictive activity effects that are present very early in the task. Finally, the use of bimanual responses could have introduced asymmetries in stimulus representation distances, for example between items assigned to different hands. Future studies could investigate representational change and predictive activity using more implicit tasks, in which participants remain unaware of temporal regularities, and with more abstract visual stimuli. In addition, while the present study employed deterministic non-adjacent regularities, more ecological contexts such as language acquisition typically involve probabilistic regularities with graded probability distributions rather than fixed patterns.

In conclusion, our work reveals that the statistical learning of non-adjacent dependencies is not a monolithic process but a dynamic interplay between two mechanisms operating on different timescales. We demonstrate that the brain first employs rapid predictive activity to guide behavior through a noisy environment, with anticipatory signals emerging early across attentional and sensorimotor networks. This initial, fast-acting mechanism is followed by a slower, more gradual representational change, which solidifies learned regularities into stable neural codes. This temporal hierarchy, supported by interactions between attentional and sensorimotor systems that filter irrelevant information, provides a mechanistic framework for how the brain extracts order from chaos, suggesting a fundamental principle whereby fast predictions pave the way for lasting representational plasticity.

Methods

All analyses were performed using the Python programming language (v3.10.9, Python Software Foundation, https://www.python.org/) and main packages including MNE-Python[37] and Scikit-learn[38].

Participants

Participants were recruited in France through promotional flyers. Fifteen healthy young adults took part in the experiment (10 women, 5 men; mean age = 25.8 ± 2.3 years). Handedness was measured by the Edinburgh Handedness Inventory[39]; the mean Laterality Quotient (LQ) of the sample varied between +33.3 and +100 (−100 means complete left-handedness, 100 means complete right-handedness; $mean_{LQ}$ = 76.0 ± 22.8). Participants had normal or corrected-to-normal vision, none of them reported a history of any neurological and/or psychiatric condition, and none of them were taking any psychoactive medication. To ensure that participants did not differ in executive function performance, the short form of the Counting Span test was administered before the magnetoencephalography (MEG) experimentation, for which they performed in the normal range (mean = 3.78 ± 0.95). All participants provided written informed consent before enrollment and were compensated a total of 40€ for the entire experiment. The experiment was conducted in accordance with the guidelines of the Declaration of Helsinki and was approved by the “Comité de Protection des Personnes Sud Méditerranée I”.

Behavioral task and experimental procedures

Statistical learning was measured with a cued version of the alternating serial reaction time (cASRT) task[40,41]. In this task, an arrow stimulus appears at the center of the screen and participants are instructed to press the key on a response device that corresponds to the spatial direction of the actual arrow as quickly and accurately as possible. Participants used their index and middle finger of both their hands to respond. The cASRT task was written, delivered to the participants, and monitored using Presentation software (v. 18.1, Neurobehavioral Systems). Visual stimuli were back-projected on a translucent screen in front of the participants. Motor responses were collected

using two LUMItouch optical response keypad with two keys each (i.e., left and up response keys on the left pad; down and right keys on the right pad). The presentation of arrow stimuli follows an alternation of pattern (P) and random (r) trials alternate with each other (e.g., 2–r–1–r–3–r–4–r; where numbers denote the four predefined spatial directions [1 = left, 2 = up, 3 = down, 4 = right] of the arrows and 'r' denote randomly chosen directions out of the four possible ones) (Fig. 1a). Arrows corresponding to pattern trials were displayed in yellow and arrows corresponding to random trials in red. Participants were told that yellow arrows always follow a predefined sequence of spatial directions while red arrows always point to a randomly chosen spatial direction out of the four possible ones. They were instructed to find the hidden predefined sequence defined by the yellow arrows to improve their performance, thus, to be faster and more accurate using this sequence information to predict the next sequence (pattern) element. However, they were not informed about the exact length of the sequence. Instructions emphasize that participants also have to respond to red (random) arrows. Out of the 24 possible permutations of the four spatial directions, 6 unique permutations were selected and assigned to each participant in a pseudorandom manner. In such a sequence, some pairs become predictable (i.e., pattern pairs) in contrast to some that stay unpredictable (i.e., random pairs) (Fig. 1b). The timing of an experimental trial was the following. The target stimulus (arrow) was presented at the center of the screen for 200 ms, then a blank screen was displayed until the participant gave a motor response but no longer than 500 ms. After a fixed delay of 750 ms following the response (indicated by a blank screen again), the next trial started (Fig. 1a). In the case of an incorrect response, an 'X' was presented at the center of the screen for 500 ms before the 750 ms blank screen delay. If no response was recorded within 700 ms following stimulus onset—the duration encompassing both stimulus presentation and the response window—a '!' appeared at the center of the screen for 500 ms, followed by a 250 ms blank screen before the onset of the next trial. Participants first performed with 3 blocks of only random trials (i.e., practice session). Then they performed 4 sessions of 5 blocks each incorporating pattern trials (i.e., learning sessions). Each block included 85 trials (for a total of 1700 trials in the entire testing session) after five warm-up trials consisting only of random stimuli. After each block, participants received a 3 s feedback on their mean reaction time (RT) and accuracy in the given block. After the feedback, explicit knowledge about the sequence was measured. Participants were instructed to continuously type the order of yellow arrows with the corresponding response keys. This sequence report lasted until participants gave 12 consecutive responses (i.e., ideally, they reported the given sequence three times). Participants were not informed that they had to provide exactly 12 responses and that the series of responses could be three repetitions of the given sequence. The explicit test was followed by a self-paced short break while participants could rest. The entire experimental procedure lasted an average of 56 minutes including the 6 min practice session and the 2 min MEG recalibration between each testing block. Structural MRI data were acquired prior to the MEG experiment from 11 participants using a Siemens Prisma 3T scanner equipped with a 64-channel coil. High-resolution 1 mm isotropic 3D magnetization-prepared rapid gradient echo (MPRAGE) T1-weighted images were obtained (repetition time = 2400 ms; echo time = 2.98 ms; matrix size = 192 × 256 × 256).

Behavioral analysis

Stimulus repetition (e.g., 222, 444) and trills (e.g., 313, 424) were excluded[42], and only correct responses were considered for further analysis. Mean RTs for pattern and random pairs were calculated separately for each participant and session. A 2 (trial conditions: pattern vs. random pairs) x 4 (sessions: 1, 2, 3, 4) two-way repeated-measures analysis of variance (ANOVA) was performed to compare mean RTs throughout the experiment. Greenhouse–Geisser correction was applied to the reported *p* values. For each session, the learning index was calculated with the following formula, and tested against 0 with a one-sample *t*-test:

$$Learning\ index\ =\ Mean\ RTs_{\ random\ pairs}\ -\ Mean\ RTs_{\ pattern\ pairs}$$

MEG recordings and preprocessing

Neuromagnetic activity was recorded with a sampling rate of 2034.5 Hz on a whole head 248 magnetometers MEG system (4D Neuroimaging magnes 3600). The MEG apparatus was housed in a magnetically shielded room. The head position was recorded before and after each block. One electro-cardiography channel (ECG) and 2 electro-oculography (EOG) channels were also recorded and realigned during preprocessing. All trials were considered irrespective of the response given by participants. Head position of the participants were realigned to that of the first learning session. We first applied a high pass filter above 1 Hz to the raw data. Independent component analysis (ICA) was then performed on this high-pass-filtered raw data to extract 30 components. To remove components linked to eye blinks/movements or heartbeats artifacts, we computed *Pearson* correlation of each ICA component time course with each of the ECG/EOG channels. ICA components that correlated above a correlation coefficient threshold of 0.9 with one of the ECG/EOG channels were discarded. The remaining ICA components

were back projected to the MEG sensors using the original raw data (unfiltered). This ICA-cleaned raw data was then bandpass filtered between 0.1 and 30 Hz. Epochs were generated at a sampling frequency of 101.73 Hz (i.e., twenty-fold decimation against raw data). For RSA analyses, epochs were created time-locked on stimulus onset with a duration of 0.8 s (i.e., from 0.2 s before the event to 0.6 s after the event) and were baselined using the 0.2 s time window preceding the visual stimulus onset. For temporal generalization and decoding analyses, epochs were generated from 4 s before and 4 s time-locked on stimulus onset, and were not baselined in order to avoid eliminating any predictive activity effects. For one participant with unusable practice session data, the first learning block was used in its place when necessary.

Source reconstruction

Participants' MEG data were co-registered to their meshed structural MRIs obtained from Freesurfer[43,44] through identification of the same anatomical landmarks (left and right pre-auricular points, and nasion). For participants (n=3) who did not undergo MRI acquisition, MEG data were morphed to a standard template brain from the FreeSurfer package. The Schaefer atlas[45] was used to parcellate the brain into 200 cortical regions. These regions were clustered into 7 different networks: the visual, sensorimotor, dorsal attention, salience, limbic, central executive, and default mode networks. Additionally, we included 3 subcortical regions -the hippocampus, the thalamus, and the cerebellum- using the Desikan-Killiany parcellation[46]. These regions served as our regions of interest (ROIs). For RSA analyses, noise covariance was computed with the 200 ms preceding stimulus onset, while data covariance was computed on post-stimulus onset data of each epoch. For temporal generalization and decoding analyses, noise covariance was computed on random trials only in the 200 ms preceding stimulus onset, while data covariance was computed on whole respective epochs. To estimate the time series in each ROI, the linearly-constrained minimum variance (LCMV) beamformer was computed on single-trial data. This particular spatial filter was chosen due to its ability to resolve deeper sources and suppress external noise on time series data[29]. Vector source estimates were used to keep the information of all (x, y, z) directions of the dipoles. We then performed Singular Value Decomposition in a time-resolved manner to only keep the most informative direction at each time point of the source estimates.

Representational similarity analysis

An Ordinary Least Squares (OLS) linear regression was applied to each sensor/vertex and to each time point to compute the model coefficients that capture the neural response patterns associated with each stimulus[47]. The model residuals from the OLS were used to estimate the covariance matrix using Ledoit-Wolf shrinkage for more stable covariance estimates[48]. The 10 fold cross-validated Mahalanobis distance was selected as the dissimilarity metric due to its sensitivity in multivariate statistical testing for RSA in neuroimaging data[26,49,50]:

$$cvMD(A,B) = (\mu_A^{train} - \mu_B^{train})^\top . \Sigma^{-1} . (\mu_A^{test} - \mu_B^{test})$$

where μ are class means estimated separately in training and testing sets, Σ is the noise covariance estimated from training residuals using Ledoit-Wolf shrinkage. This resulted in time-resolved representational dissimilarity matrices (RDMs) for each participant and each session. For each participant, the distance between elements of pattern and random pairs was extracted from the RDMs. Then the similarity index was computed using the following formula:

$$Similarity\ index = cvMD\ learning_{random} - cvMD\ learning_{pattern}$$

This index reflects changes in similarity with pattern pairs relative to changes in similarity within random pairs. A positive value for the similarity index means that the similarity between elements within pattern pairs is higher than the similarity among elements within random pairs. To test if the similarity index was significantly different than 0 at the group level, a non-parametric *t* test corrected for multiple comparisons with cluster-based permutations was performed, using the default parameters of MNE-Python. Spearman's correlation was also computed in a time-resolved manner between the similarity index and the learning index of each participant. The resulting rhos for each participant were tested against 0 using a one-sample *t* test. This pipeline was applied both in sensor space (whole brain) and source space (ROIs).

Temporal generalization

Except for correlation analyses, only learning session trials were considered in order to prevent dampening of found effects by the practice session. Epochs were generated time-locked -4 s before and 4 s after stimulus onset. A classifier that has been trained at a specific time sample was tested at all possible time samples using the temporal generalization method[30], resulting in two-dimensional arrays of decoding performance in accuracy (i.e., generalization matrix) for each participant at each session. This was done for both pattern and random trials. To test if the group-mean accuracy of the decoding performance is above chance level (25%) at each time point, a non-parametric *t* test corrected for multiple comparisons with cluster-based permutations was computed[51]. Next, the contrast was computed by subtracting the random generalization matrix to the pattern one to evidence any predictive activity prior to stimulus onset. To test if this difference in accuracy was significant, we performed a non-parametric *t* test corrected for multiple comparisons with cluster-based permutations[51]. This pipeline was applied for pattern and random trials in sensor space (whole brain) only.

Decoding

Except for correlation analyses, only learning session trials were considered in order to prevent dampening of found effects by the practice session. Multivariate pattern analysis (MVPA) decoding aimed at predicting the value of a specific variable *y* (i.e., the spatial arrows) from the brain signal *X*. The multiclass Logistic Regression was selected as the classifier. The data (*X*) were whitened using a standard scaler that *z* scored each channel at each time point across trials. Each estimator was fitted on each participant separately, across all MEG sensors/source vertices, and at a unique time sample, using the default parameters of scikit-learn. The cross-validation was performed using a 10-fold stratified folding, and accuracy was selected as the metric used to assess classifier's performance. To test if the group-mean accuracy of the decoding performance is above chance level (25%) at each time point, a non-parametric *t* test corrected for multiple comparisons with cluster-based permutations was computed[51]. Epochs of each session (including practice) were concatenated and used as a single epoch to maximize the number of trials in the training and testing set. This pipeline was applied for pattern and random trials in source space (ROIs) only.

Non-linear temporal modeling

The temporal evolution of behavioral performance, predictive activity and representational similarity over the time course of the experiment was investigated using generalized additive mixed models (GAMM) using the mgcv[52] package (version 1.9.0) in R (version 4.3.2). Each variable was modelled with a GAMM including a single fixed effect: a smooth penalized regression spline of *block*. Therefore, *block* was treated as a continuous predictor with 23 values (3 practice blocks + 20 task blocks). Participant heterogeneity was captured by a random intercept and a random linear slope, allowing individual temporal trajectories to deviate from the population curve both by a constant offset and a linear trend. In order to formally compare temporal dynamics across the three metrics of interest, a fourth model was fitted to the standardized values. Each metric was scaled to have a unit standard deviation at the group level. The three metrics were grouped into a single factor variable which was allowed to interact with *block* via the addition of a factor smooth interaction in the model predictors. This way, each (standardized) metric is modeled by its deviation from the mean of the three, similar to a sum-to-zero contrast (*aka* sum contrast coding) in a linear model. A significant overall test for this interaction term indicates that at least one metric's temporal trajectory differs from the others. Submodels including only two of the three metrics were fitted as *post hoc* tests to identify the differing pairs. All models used thin-plate regression splines and a basis dimension of 10 (the default values), and assumed Gaussian errors with an identity link. The adequacy of these parameters and validity of the Gaussian assumption was inspected with *ad hoc* diagnostics provided by mgcv. In particular, for all models residuals were found to be normally distributed and estimated degrees-of-freedom of fixed effect smooths were well below 10. Parameters and smooth penalties were estimated by restricted maximum likelihood (REML, the default).

References

1. Pedraza, F. *et al.* Evidence for a competitive relationship between executive functions and statistical learning. *Npj Sci. Learn.* **9**, 30 (2024).

2. Schapiro, A. & Turk-Browne, N. Statistical Learning. in *Brain Mapping* 501–506 (Elsevier, 2015). doi:10.1016/B978-0-12-397025-1.00276-1.

3. Saffran, J. R., Aslin, R. N. & Newport, E. L. Statistical Learning by 8-Month-Old Infants. *Science* **274**, 1926–1928 (1996).

4. Gómez, R. L. Variability and Detection of Invariant Structure. *Psychol. Sci.* (2002).

5. Wilson, B. *et al.* Non-adjacent Dependency Learning in Humans and Other Animals. *Top. Cogn. Sci.* **12**, 843–858 (2020).

6. Christiansen, M. H. & Chater, N. The language faculty that wasn't: a usage-based account of natural language recursion. *Front. Psychol.* **6**, (2015).

7. Malassis, R., Rey, A. & Fagot, J. Non-adjacent Dependencies Processing in Human and Non-human Primates. *Cogn. Sci.* **42**, 1677–1699 (2018).

8. Wilson, B., Smith, K. & Petkov, C. I. Mixed-complexity artificial grammar learning in humans and macaque monkeys: evaluating learning strategies. *Eur. J. Neurosci.* **41**, 568–578 (2015).

9. Gómez, R. & Maye, J. The Developmental Trajectory of Nonadjacent Dependency Learning. *Infancy* **7**, 183–206 (2005).

10. Conway, C. M. *et al.* Distinct neural networks for detecting violations of adjacent versus nonadjacent sequential dependencies: An fMRI study. *Neurobiol. Learn. Mem.* **169**, 107175 (2020).

11. Kóbor, A. *et al.* Finding Pattern in the Noise: Persistent Implicit Statistical Knowledge Impacts the Processing of Unpredictable Stimuli. *J. Cogn. Neurosci.* **36**, 1239–1264 (2024).

12. Henin, S. *et al.* Learning hierarchical sequence representations across human cortex and hippocampus. *Sci. Adv.* (2021).

13. Schapiro, A. C., Kustner, L. V. & Turk-Browne, N. B. Shaping of Object

Representations in the Human Medial Temporal Lobe Based on Temporal Regularities. *Curr. Biol.* **22**, 1622–1627 (2012).

14. Greco, A., Moser, J., Preissl, H. & Siegel, M. Predictive learning shapes the representational geometry of the human brain. *Nat. Commun.* **15**, 9670 (2024).

15. Schapiro, A. C., Rogers, T. T., Cordova, N. I., Turk-Browne, N. B. & Botvinick, M. M. Neural representations of events arise from temporal community structure. *Nat. Neurosci.* **16**, 486–492 (2013).

16. Greco, A., Moser, J., Preissl, H. & Siegel, M. Predictive learning shapes the representational geometry of the human brain. Preprint at https://doi.org/10.1101/2024.03.07.583842 (2024).

17. Miyashita, Y. Neuronal correlate of visual associative long-term memory in the primate temporal cortex. *Nature* **335**, 817–820 (1988).

18. Karuza, E. A., Emberson, L. L. & Aslin, R. N. Combining fMRI and behavioral measures to examine the process of human learning. *Neurobiol. Learn. Mem.* **109**, 193–206 (2014).

19. Wacongne, C. *et al.* Evidence for a hierarchy of predictions and prediction errors in human cortex. *Proc. Natl. Acad. Sci.* **108**, 20754–20759 (2011).

20. De Lange, F. P., Heilbron, M. & Kok, P. How Do Expectations Shape Perception? *Trends Cogn. Sci.* **22**, 764–779 (2018).

21. Friston, K. The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* **11**, 127–138 (2010).

22. Stefanics, G., Heinzle, J., Horváth, A. A. & Stephan, K. E. Visual Mismatch and Predictive Coding: A Computational Single-Trial ERP Study. *J. Neurosci.* **38**, 4020–4030 (2018).

23. Park, S.-W. *et al.* Developmental change in predictive motor abilities. *iScience* **26**, 106038 (2023).

24. Kok, P., Jehee, J. F. M. & de Lange, F. P. Less Is More: Expectation Sharpens Representations in the Primary Visual Cortex. *Neuron* **75**, 265–270 (2012).

25. Demarchi, G., Sanchez, G. & Weisz, N. Automatic and feature-specific prediction-related neural activity in the human auditory system. *Nat. Commun.* **10**, 3440 (2019).

26. Geun Kim, M. Multivariate outliers and decompositions of mahalanobis distance. *Commun. Stat. - Theory Methods* **29**, 1511–1526 (2000).

27. Diedrichsen, J., Provost, S. & Zareamoghaddam, H. On the distribution of cross-validated Mahalanobis distances. Preprint at https://doi.org/10.48550/arXiv.1607.01371 (2016).

28. Van Veen, B. D., Van Drongelen, W., Yuchtman, M. & Suzuki, A. Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. *IEEE Trans. Biomed. Eng.* **44**, 867–880 (1997).

29. Westner, B. U. *et al.* A unified view on beamformers for M/EEG source reconstruction. *NeuroImage* **246**, 118789 (2022).

30. King, J.-R. & Dehaene, S. Characterizing the dynamics of mental representations: the temporal generalization method. *Trends Cogn. Sci.* **18**, 203–210 (2014).

31. Bago, B. & De Neys, W. Fast logic?: Examining the time course assumption of dual process theory. *Cognition* **158**, 90–109 (2017).

32. Katsuki, F. & Constantinidis, C. Bottom-Up and Top-Down Attention: Different Processes and Overlapping Neural Systems. *The Neuroscientist* **20**, 509–521 (2014).

33. Vossel, S., Geng, J. J. & Fink, G. R. Dorsal and Ventral Attention Systems: Distinct Neural Circuits but Collaborative Roles. *The Neuroscientist* **20**, 150–159 (2014).

34. Macaluso, E. Orienting of spatial attention and the interplay between the senses. *Cortex* **46**, 282–297 (2010).

35. Park, J., Janacsek, K., Nemeth, D. & Jeon, H.-A. Reduced functional connectivity supports statistical learning of temporally distributed regularities. *NeuroImage* **260**, 119459 (2022).

36. Vékony, T. *et al.* Modality-specific and modality-independent neural representations work in concert in predictive processes during sequence learning. *Cereb. Cortex* **33**, 7783–7796 (2023).

37. Gramfort, A. *et al.* MNE software for processing MEG and EEG data. *NeuroImage* **86**, 446–460 (2014).

38. Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *Mach. Learn. PYTHON* (2011).

39. Oldfield, R. C. The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia* **9**, 97–113 (1971).

40. Howard, J. H. & Howard, D. V. Age Differences in Implicit Learning of Higher Order Dependencies in Serial Patterns. *Psychol. Aging* **12**, 23 (1997).

41. Nemeth, D., Janacsek, K. & Fiser, J. Age-dependent and coordinated shift in performance between implicit and explicit skill learning. *Front. Comput. Neurosci.* **7**, (2013).

42. Ambrus, G. G. *et al.* When less is more: Enhanced statistical learning of non-adjacent dependencies after disruption of bilateral DLPFC. *J. Mem. Lang.* **114**, 104144 (2020).

43. Dale, A. M., Fischl, B. & Sereno, M. I. Cortical Surface-Based Analysis.

44. Fischl, B. FreeSurfer. *NeuroImage* **62**, 774–781 (2012).

45. Schaefer, A. *et al.* Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic Functional Connectivity MRI. *Cereb. Cortex* **28**, 3095–3114 (2018).

46. Desikan, R. S. *et al.* An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage* **31**, 968–980 (2006).

47. Grootswagers, T., Wardle, S. G. & Carlson, T. A. Decoding Dynamic Brain Patterns from Evoked Responses: A Tutorial on Multivariate Pattern Analysis Applied to Time Series Neuroimaging Data. *J. Cogn. Neurosci.* **29**, 677–697 (2017).

48. Ledoit, O. & Wolf, M. A well-conditioned estimator for large-

dimensional covariance matrices. *J. Multivar. Anal.* **88**, 365–411 (2004).

49. Walther, A. *et al.* Reliability of dissimilarity measures for multi-voxel pattern analysis. *NeuroImage* **137**, 188–200 (2016).

50. Guggenmos, M., Sterzer, P. & Cichy, R. M. Multivariate pattern analysis for MEG: A comparison of dissimilarity measures. *NeuroImage* **173**, 434–447 (2018).

51. Maris, E. & Oostenveld, R. Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci. Methods* **164**, 177–190 (2007).

52. Wood, S. N. *Generalized Additive Models: An Introduction with R*. (Chapman and Hall/CRC, 2017). doi:10.1201/9781315370279.

Data Availability

Raw MEG data to reproduce all the results will be released in an openly available repository upon publication.

Code Availability

Data analysis was carried out with Python and R. Code will be released in an openly available repository upon publication.

Acknowledgements

This work was performed within the framework of the LABEX CORTEX(ANR-11-LABX-0042) of l'Université Claude Bernard Lyon 1, within the program "Investissements d'Avenir" (decision n° 2019-ANR-LABX-02) operated by the French National Research Agency (ANR). This work was also supported by the French National Grant Agency (ANR-22-CPJ1-0042-01 and ANR-24-CE37-5807) (to D.N.); the National Brain Research Program project NAP2022-I-2/2022 (to D.N.). We are grateful to Arnaud Rey, whose ideas and inspiring conversations were instrumental in shaping and realizing this project, and to Denis Schwartz for his valuable insights and expertise on source reconstruction.

Author contributions

C.T.: Conceptualization, Formal analysis, Investigation, Methodology, Visualization, Writing—original draft, Writing—review & editing.
T.V.: Conceptualization, Software, Data curation, Writing—review & editing.
O.A.: Formal analysis, Investigation, Methodology, Visualization, Writing—original draft, Writing—review & editing.
L.T.: Data curation, Writing—review & editing.
A.B.: Writing—review & editing.

M.V.: Writing—review & editing.
D.N.: Conceptualization, Funding acquisition, Methodology, Supervision, Project administration, Writing—original draft, Writing—review & editing.
R.Q.: Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Supervision, Project administration, Writing—original draft, Writing—review & editing.

Fundings

This work was supported by the ATIP-Avenir program.

Competing interests

The authors declare no competing interests.