Human statistical learning dynamically shapes the hippocampal processing of temporal associations

Wei Tang[1, *], Morten H. Christiansen[2,3], Zhenghan Qi[4]

[1] Department of Computer Science and Department of Psychological and Brain Sciences, Indiana University-Bloomington, Bloomington, IN 47408

[2] Department of Psychology, Cornell University, Ithaca, NY 14853

[3] Haskins Laboratories, New Haven, CT 06511

[4] Department of Communication Sciences and Disorders and Department of Psychology, Northeastern University, Boston, MA 20115

* Corresponding author, wt1@iu.edu

Abstract

Recent research suggests that the human hippocampus becomes sensitive to temporal associative patterns during statistical learning (SL). And yet, how such sensitivity develops throughout the course of learning remains unclear. We investigated this question in human adults performing a visual SL task, in which the transition probability among stimuli varied systematically across learning sessions. Hippocampal activity acutely reflected this cross-session change and was predictive of behavioral performance. Using a hidden Markov model to decode stimulus patterns from hippocampal activity, we observed a "mere-exposure" effect, such that brief experience with the temporally structured stimuli was sufficient to alter the hippocampal processing. As learning unfolded, feedback from the repeated exposure shaped the hippocampal processing to better accommodate the environmental statistics. Our findings provide direct

evidence of how experience with associative patterns sculpts hippocampal activity in a nonlinear fashion, bringing new insight to its active role in SL.

Introduction

Humans are able to learn patterned regularities in the environment based on the co-occurrence of events in the spatial and temporal domains, an ability called statistical learning (SL). In the temporal domain, participants in an SL task become sensitive to the associative relationships between sequentially presented stimuli (Fiser & Aslin, 2002; Saffran et al., 1996; Turk-Browne et al., 2005). Such temporal association can be characterized (at least in part) by the transition probability (TP) between events, indicating how likely one stimulus is followed by another. Behaviorally, high TP facilitates on-line sensory processing and later memory measurements of the learned events. For example, the reaction time (RT) to probes during learning is often shorter for stimuli with high TP than those with low TP (Batterink, 2017; Schneider et al., 2020). Post-learning memory tests show that higher accuracy is associated with items that were presented with a statistically patterned temporal structure during study, compared with those in random order (Isbilen et al., 2020; Turk-Browne et al., 2005).

Recent neuroimaging studies provide preliminary evidence for a multi-level framework across cognitive and sensory domains underlying the SL processes (Conway, 2020). At the sensory level, statistical regularity facilitates modality-specific processing, inducing short-term plasticity in early visual and auditory regions (Karuza et al., 2013; McNealy et al., 2006; Turk-Browne et al., 2009, 2010). At the cognitive level, higher functions such as long-term memory, working memory, and selective attention help develop sensitivity to statistical regularities beyond sensory features. For example, both linguistic and nonlinguistic learning activates

associative prefrontal regions (Bischoff-Grethe et al., 2000; Furl et al., 2011; Karuza et al., 2013). To achieve the learning goals, the brain network engages both domain-general and modality-specific processing, allowing expectations to be generated from learned regularities to facilitate perception (Conway, 2020; Reber, 2013; Thothathiri & Rattinger, 2015; Frost et al., 2015).

The multi-level processing framework is yet to be extended to account for hippocampal activity in SL. Hippocampal neurons code for association among objects/events that occur close in time and space (Burgess et al., 2002; Ekstrom & Ranganath, 2018; O'Reilly & Norman, 2002; Schiller et al., 2015). These conjunctive representations are insensitive to specific visual or auditory features. Moreover, hippocampal activations reorganize in response to temporally associated stimuli after only minutes of exposure (Schapiro et al., 2012). Thus, an intriguing possibility is that hippocampal neurons may play an important role in the rapid development of sensitivity to statistical regularities (Schapiro et al., 2017; Turk-Browne, 2019). The learned sensitivity to statistical regularities can then further serve additional functions such as context-based memory retrieval or prediction of future states/rewards (Sherman & Turk-Browne, 2020).

Nevertheless, the formation of a neural representation does not necessitate an active role in learning *per se*. For example, conjunctive representations can be stored in the hippocampus as a learning *outcome*, while the learning *process* is carried out elsewhere in the brain. Indeed, some have argued that an effective long-term memory storage is critical for successful SL (Batterink & Paller, 2017). In contrast, Schapiro and colleagues (Schapiro et al., 2017) used biologically inspired computational models to show that circuits of hippocampal subfields can actively track and update TP, whose activity accounts for the observed representational similarity. However, despite the model predictions, experimental evidence of this rapid learning process has been

scarce, and the exact role of hippocampus in SL remains an open question (Covington et al., 2018; Schapiro et al., 2014).

In the current study, we investigated whether the human hippocampus actively tracks and changes its processing of temporal regularity in the course of visual SL. If the hippocampus plays an active role in learning, we expect an acute change in activity in response to the change of structure/randomness in the stimuli. Moreover, SL-related brain activity is expected to reflect SL-related behavior. To assess the SL effect, we measured the similarity of hippocampal blood-oxygenation-level-dependent (BOLD) activity during the presentations of neighboring stimuli (i.e., lag-1 autocorrelation) and examined the change of this similarity over the course of learning. Our approach was inspired by previous functional magnetic resonance imaging (fMRI) work showing that multi-voxel BOLD activity exhibited similar spatial patterns in response to stimuli that co-occurred in time (Schapiro et al., 2012, 2013). This multi-voxel similarity implies lag-1 BOLD autocorrelation in individual voxels. Importantly, the autocorrelation measure allowed us to build formal models to link hippocampal activity with stimulus transition patterns (as we will describe in detail below). Specifically, we asked whether this autocorrelation pattern would 1) be sensitive to the TP difference between structured and random stimuli, 2) reorganize to reflect the switch between structure and randomness in the stimuli, and 3) correlate with SL-related behavior. Although different factors can contribute to BOLD autocorrelation, including the shape of the hemodynamic response function, only SL-related factors would drive the change of autocorrelation to reflect the change of temporal patterns among stimuli. To further delineate the hippocampal sensitivity to specific temporal statistics, we used a Hidden Markov Model (HMM) to decode stimulus transition patterns from the autocorrelated hippocampal activity. In the HMM, we assumed that the level of BOLD autocorrelation depended on the stimuli's

positions in a structured sequence rather than their sensory features. By examining the decoding efficacy at different learning stages, we explored how hippocampal sensitivity to TP was shaped by repeated exposure to structured and random sequences. We further examined whether the BOLD autocorrelation and decoding patterns differed among hippocampal subfields. The results suggest that the dynamic process of learning shapes the hippocampal representations of statistical temporal patterns, bringing new insight into the active role of the hippocampus in visual SL.

Results

Learning effects in behavior and hippocampal activity. Subjects viewed sequentially presented images while responding to target images embedded in each sequence (Fig. 1). The sequence blocks varied along two dimensions: the stimuli in each block were (1) linguistic or non-linguistic and (2) temporally arranged into triplets (S-block) or presented in random order (R-block) (see Online Methods, *stimuli*). The temporal structure was set to systematically differ between linguistic and nonlinguistic stimuli across two learning sessions (Fig. 1A). In Session 1, linguistic S-blocks were pseudo-randomly intermixed with non-linguistic R-blocks. In Session 2, the structure/randomness was switched, such that linguistic R-blocks were pseudo-randomly intermixed with nonlinguistic S-blocks.

The target was set to be the third element of a randomly selected triplet from either the linguistic or nonlinguistic sequences, and its location followed no systematic pattern in each block either within- or across-participants (see Online Methods, *procedure*). Subjects were not informed of the temporal structure and completed the target-detection task with an average hit rate of 97.4 ± 4.7% (cross-subject median ± s.d.), false-alarm rate of 0.2 ± 0.2%, and reaction time (RT) of 446 ± 52 ms. Comparisons of RT between S- and R-blocks showed a facilitation

effect on target detection. For both linguistic and nonlinguistic stimuli, across subjects, the trial-average RT was significantly lower in S-blocks than in R-blocks (Fig. 1B; linguistic: Wilcoxon signed-rank $W = 23.0$, $p < 0.00028$; nonlinguistic: $W = 71.0$, $p < 0.024$). Correspondingly, for both types of stimuli and across hippocampal voxels (Fig. 1C & D), lag-1 autocorrelation of BOLD activity was significantly greater in S- than in R-blocks (Fig. 1D; linguistic: paired-sample $t_{11262} = 83.58$, $p < 1\times 10^{-6}$; nonlinguistic: $t_{11262} = 77.12$, $p < 1\times 10^{-6}$).

Temporal evolution of RT and BOLD autocorrelation during learning. We examined the change of RT and BOLD autocorrelation with respect to the change of temporal structure among stimuli. By design, within each session, TP differed between S- and R-blocks; across sessions, there was a switch between structure/randomness for each type of stimuli. Accordingly, the variation of RT across mini blocks depended on the session and block type (Fig. 2A, within-subject, block × session × block type interaction, $F_{1, 452.99} = 4.84$, $p < 0.028$). The cross-session RT change for each type of stimuli can account for this three-way interaction. Linguistic stimuli were presented as triplets in Session 1 (S-blocks, numbered 1-6 in Fig. 2A) and randomly in Session 2 (R-blocks, numbered 7-12). Correspondingly, within-subject, RT significantly increased from Session 1 to Session 2 (Wilcoxon test for median RT between-session, $W = 9.0$, $p < 0.00033$). In contrast, Nonlinguistic stimuli were presented randomly in Session 1 (R-blocks, numbered 1-6) and then in triplets in Session 2 (S-blocks, numbered 7-12). RT significantly decreased from Session 1 to Session 2 within-subject (Wilcoxon test for median RT between-session, $W = 30.0$, $p < 0.0051$). Indeed, the cross-subject median RT continuously decreased across concatenated mini blocks (slope $b = -4.27$, $p < 0.0059$).

The BOLD autocorrelation across mini blocks also varied depending on the session and block type (Fig. 2A, within-subject, block × session × block type interaction, $F_{1, 434.00} = 12.83$, p

138 < 0.00038). Although visual inspection suggested cross-session differences, the same post-hoc
139 comparisons as those in the RT analysis were not significant. This was not surprising because
140 autocorrelation could be affected by a variety of physiological factors that were unrelated to the
141 task. For example, intrinsic brain dynamics during the resting state could also lead to lag-1
142 autocorrelation in the BOLD activity (Huang et al., 2018). These dynamics would not change
143 according to environmental statistics, and thus, could contribute to the insignificance of cross-
144 session comparisons of the autocorrelation. Alternatively, the contrast between S- and R-blocks
145 would eliminate the common factors and leave those that were sensitive to the temporal
146 statistical difference among stimuli.

147 Thus, as an alternative post-hoc test to break down the three-way interactions found by
148 the ANOVA, we examined the S-R contrast of BOLD autocorrelation across mini blocks. In
149 each session, nonlinguistic blocks were subtracted from linguistic blocks, resulting in an S-R
150 difference for Session 1 and R-S difference for Session 2 (Fig. 2B). This subtraction order was
151 kept to facilitate the detection of cross-session differences: if the BOLD autocorrelation rapidly
152 adapted to the new temporal statistics, we would expect a sign-flip in Session 2. Indeed, the sign-
153 flip was evident in both the RT and BOLD autocorrelation measurements (Fig. 2B, within-
154 subject main effect of session, RT: $F_{1, 19} = 21.44$, $p < 0.00018$; BOLD autocorrelation: $F_{1, 82.30} =$
155 9.00, $p < 0.0035$). Across individual blocks, the hippocampal and behavioral measurements were
156 correlated, such that the S-R contrast of BOLD autocorrelation was linearly proportional to that
157 of RT (Fig. 2C, Spearman's $r = -0.62$, $p < 0.035$).

158 *HMM decoding with simulated data.* The BOLD autocorrelation measures the overall
159 similarity of the signal at neighboring points of a sequence. By design it does not resolve
160 temporal structures within a sequence. We were interested in the hippocampal sensitivity to

161 triplet boundaries in addition to its overall sensitivity to structure versus randomness. Therefore,
162 we devised an HMM approach to further test whether TP information can be decoded from the
163 hippocampal activity (Fig. 3A). The conditional-Gaussian HMM designed for this study
164 contained three hidden states: *start*, *mid* and *end*, corresponding to the three positions in a triplet.
165 Importantly, the model distinguishes stimuli by their positions but not the specific visual content.
166 If the BOLD autocorrelation differed with respect to stimulus position (e.g. within-triplet vs.
167 between-triplets), the HMM decoding would yield a triplet transition matrix (Fig. 3A). For the
168 random sequences, there was no inherent *start-mid-end* structure, and the BOLD autocorrelation
169 was assumed to be uniform among all stimuli. The corresponding decoding result was expected
170 to randomly deviate from the triplet transition matrix.

171 We first tested the model's decoding efficacy with simulated data (Fig. 3B-D), in which
172 the hidden states were strictly ordered into *start-mid-end* triplets. Theoretically, the decoding
173 accuracy is affected by two parameters: the sensitivity index d' and the autocorrelation ρ. The
174 sensitivity index indicates how well the observed BOLD data can be separated by their mean
175 values with respect to the hidden states. For example, given mean values of 0.4, 0.5 and 0.6
176 corresponding to *start, mid* and *end*, with an equal standard deviation of 0.1, the d' of this
177 pseudo-BOLD data equals 1.3 (see Eq. 11 in Online Methods) and indicates a good separation.
178 The pattern of BOLD autocorrelation ρ determines the extent to which the BOLD data is
179 indicative of the underlying temporal structure among states. For example, given ρ of 0.8 for
180 state pairs *start-mid* and *mid-end*, 0.2 for *end-start*, and 0 for any other pair, this autocorrelation
181 pattern represents high temporal association within a triplet, low temporal association across
182 triplets, and zero association for implausible state transitions. When the observed data are
183 generated with a high sensitivity index and triplet-representing ρ values, the HMM decoding will

yield a triplet transition matrix among hidden states. To demonstrate the model efficacy, we set *d'* and ρ to the values described above and generated surrogate observations (Fig. 3B). We compared the decoded transition matrix against the ground truth using normalized mean-squared error (n-MSE, Eq. 10 in Online Methods), which grows proportionally with the deviation from the ground truth. The HMM successfully decoded the transition matrix with 100% accuracy (n-MSE = 0) in each of the 100 runs.

This simulated example represented an ideal case for decoding. However, the parameters used here may differ from those of the veridical BOLD data. We next conducted numerical experiments on different parameter settings to test how they affected decoding accuracy (Fig. 3C & D). First, we varied the sensitivity index while keeping ρ fixed to the ideal values. When *d'* dropped below 0.7, the decoding became imperfect (n-MSE > 0) (Fig. 3C, scatter plot), such that nonzero probabilities were assigned to invalid transitions (i.e. false alarms). Nonetheless, even with false alarms, the decoded transition matrix was qualitatively consistent with the ground truth, such that the cells corresponding to the valid transitions had the highest probability (Fig. 3C, matrix plot). We then fixed the sensitivity index to the average value of the veridical BOLD data (0.23 ± 0.11 across subjects, marked as shaded area in Fig. 3C). We tested the decoding accuracy by varying the ρ values (Fig. 3D, scatter plot). The decoding accuracy decreased when ρ decreased, but the decoded transition matrix remained qualitatively accurate at each ρ level (Fig. 3D, matrix plot). The simulation results suggested that if data autocorrelation encoded temporal association, and given the parameter ranges of veridical BOLD data (mean ± s.d., shaded areas in Fig. 3C & D), the HMM decoding would be imperfect but remain qualitatively consistent with the true transition patterns. In the following text, for brevity, we refer to decoded

206 matrices that were qualitatively similar to the triplet-transition matrix as "triplet-like", and those
207 randomly deviating from it as "uninformative".

208 *HMM decoding with BOLD data.* We applied the HMM approach to decode transition
209 patterns in the veridical BOLD data, in each block throughout the task. As a cautionary step, we
210 validated the assumption that the BOLD autocorrelation reflected hippocampal sensitivity to TP
211 structure. We tested this assumption by comparing the conditional-Gaussian HMM, which uses
212 both the mean and autocorrelation values for decoding, with a standard univariate-Gaussian
213 HMM, which uses only the mean values for decoding (see Online Methods, *Hidden Markov*
214 *modeling*). Both analyses were performed on the voxel-averaged hippocampal BOLD time series
215 from the S-blocks in each subject. Across subjects, the conditional-Gaussian HMM showed
216 significantly greater accuracy than the univariate-Gaussian HMM (Fig. 4A, Wilcoxon $W = 0.0$, p
217 $< 8.8 \times 10^{-5}$), demonstrating the improvement of decoding using autocorrelation information.
218 Thus, the following analyses were focused on the conditional-Gaussian HMM.

219 Averaged across subjects and sessions, the decoding accuracy significantly varied across
220 S-blocks (one-way ANOVA block effect, $F_{5,114} = 2.98$, $p < 0.014$) but not R-blocks (one-way
221 ANOVA, main effect of block, $F_{5,114} = 0.92$, $p < 0.43$). Curiously, the highest accuracy occurred
222 in the early blocks (Fig. 4B). Decoded patterns of the transition matrix confirmed this
223 observation (Fig. 4C): triplet-like transition pattern was found in the second and last S-blocks, as
224 well as the first two R-blocks. Importantly, the decoded pattern was triplet-like for the last S-
225 block and noninformative for the last R-block. The triplet-like pattern in early S- and R-blocks
226 suggested that learning occurred rapidly and might be history-dependent. We further examined
227 this possibility. The rationale was that at the beginning of the task, the hippocampus might not be
228 sensitive to the S-R difference. The exposure to S-block(s) could have caused a mere-exposure

229 effect, such that the hippocampus responded to the early R-blocks as if there was a triplet
230 structure. If this was true, we should find triplet-like R-block decoding patterns exclusively in the
231 subset that was presented after S-blocks. We examined this possibility with data from Session 1.
232 Among the twenty subjects, ten (Group 1) started with an S-block and ten (Group 2) started with
233 an R-block (Fig. 4D). Separately for each group, we examined the decoded transition matrix of
234 the first block (outlined in black in Fig. 4D) as well as the following S- and R-blocks (outlined in
235 magenta and blue, respectively). The decoding outcome showed a presentation order effect (Fig.
236 4D), such that the decoded pattern was triplet-like if the first block was structured, and
237 uninformative if random. Moreover, the S- and R-blocks that followed the firstly presented S-
238 block both showed triplet-like decoded patterns. In contrast, the S- and R-blocks following the
239 firstly presented R-block showed uninformative decoded patterns.

240 *BOLD autocorrelation and decoding accuracy across hippocampal subfields.* We
241 compared the autocorrelation and decoding results among the hippocampal subfields (Fig. 5A).
242 Averaged across sessions and blocks, the greatest S-R difference of BOLD autocorrelation was
243 observed in subfields CA2/3 and CA4/DG (Fig. 5B, one-way ANOVA main effect of subfield,
244 $F_{4, 95} = 2.75$, $p < 0.032$). The decoding accuracy for S-blocks did not significantly differ across
245 subfields (one-way ANOVA on the n-MSE, main effect of subfield, $F_{4, 95} = 1.06$, $p < 0.38$).
246 Consistently, the decoded transition matrices showed similarly uninformative patterns for each
247 subfield (Fig. 5B). We further examined the autocorrelation and decoding results in the first and
248 last blocks, averaged across sessions (Fig. 5C & D). For the first block, the S-R difference of
249 BOLD autocorrelation did not significantly differ across subfields (one-way ANOVA main
250 effect of subfield, $F_{4, 95} = 0.58$, $p < 0.68$), while a triplet-like transition matrix was decoded in
251 CA1 from the S-block (Fig. 5C). For the last block, the greatest S-R difference of BOLD

252 autocorrelation was found in the parasubiculum (one-way ANOVA, near-significant main effect
253 of subfield, $F_{1, 48} = 3.59$, $p < 0.064$), in which a triplet-like transition matrix was decoded from
254 the S-block (Fig. 5D).
255
256 **Discussion**
257 This study investigates how processing in the human hippocampus changes to
258 accommodate temporal transitions of sequentially presented visual stimuli during an SL task. We
259 found that hippocampal activity is sensitive to the temporal structures in the visual inputs,
260 adaptive to the changes of the embedded temporal patterns, and rapidly alters processing as
261 reflected by a mere-exposure effect. Hippocampal sensitivity to temporal structure was reflected
262 in the autocorrelation of the hippocampal BOLD activity. We observed increasing
263 autocorrelation of hippocampal BOLD activity during the exposure to structured versus
264 randomly ordered stimuli. When the structure/randomness flipped across learning sessions, the
265 autocorrelation pattern swiftly reflected this change. Importantly, the autocorrelation pattern was
266 correlated with the RT pattern across mini blocks throughout the task, demonstrating a link
267 between the SL-related hippocampal activity and the SL-related behavior. Moreover, we found
268 that the triplet patterning can be decoded from the hippocampal activity, and that the decoding
269 outcome was history-dependent. At the beginning of the task, the decoding results depended on
270 the sequential order of S- and R-blocks being presented; at the end of the task, the decoding
271 results accurately separated the S- and R-blocks. Overall, these results suggest that the
272 hippocampus is not a passive storage for learning outcomes but plays an active role in the
273 learning process.

A long-standing theory is that the hippocampus codes for associations between objects/events that occur close in time or space (Burgess et al., 2002; Cohen & Eichenbaum, 1995; Ekstrom & Ranganath, 2018; O'Keefe et al., 1978). In accordance with previous work, we found that hippocampal activity became self-similar (measured by lag-1 autocorrelation) when the presented items were highly associated in time (Fig. 1D). Moreover, our findings demonstrate a direct link between hippocampal processing and the triplet patterning of the stimuli. To learn a statistical pattern, in addition to detecting the co-occurring items, the brain also needs to estimate how often such co-occurrence happens. In this study, this frequency information is reflected by TP, which equals 1 within a triplet, 0.25 between triplets, and 1/12 among randomly ordered stimuli. The hippocampus is sensitive to the general difference of TP between structured and random sequences, demonstrated by the S-R difference in the BOLD autocorrelation.

However, the autocorrelation measurement cannot resolve TP differences within a sequence, which critically determines the patterning of a triplet structure. Thus, we designed the HMM analysis to test the hippocampal sensitivity to triplet patterning. The conditional-Gaussian HMM assumes that the similarity of BOLD activity for neighboring stimuli can be differentiated by their locations (*start*, *mid*, or *end* position of a triplet) but not necessarily the specific visual content (linguistic or nonlinguistic). Thus, successful HMM decoding requires the BOLD autocorrelation of items within a triplet to be different from that of items across triplets. The simulation results validated this assumption by showing that when a hidden sequence generated data with said autocorrelation structure (Fig. 3B), the transition matrix was correctly decoded with 100% accuracy. Under noisy conditions that approximated true BOLD data, the decoded patterns were robustly triplet-like (Fig. 3C & D). Thus, the successful decoding in veridical

297 BOLD time series was evidence that the hippocampal activity was sensitive to the triplet TP
298 pattern within-sequence (Fig. 4C & D).

299 Importantly, the HMM decoding results suggest a proactive learning process in which the
300 current sensitivity to structure depends on prior experience. The history-dependence is implied in
301 the decoding result of early R-blocks. At the beginning of the task, if one or more S-blocks
302 preceded the first R-block, the decoded R-block transition matrix contained triplet information
303 (Fig. 4D, subject group 1). In contrast, in R-blocks that did not follow any S-block, as well as the
304 first S-block that followed R-block exposure, no triplet transition pattern was decoded (Fig. 4D,
305 subject group 2). This result indicates that initially presented stimuli may shape the processing of
306 the subsequent state of the environment. This is in broad accordance with theories of context
307 dependency and predictive computations in the hippocampus, such as successor representations
308 in reinforcement learning (Dayan, 1993; Stachenfeld et al., 2017), temporal context models in
309 episodic memory (Howard & Kahana, 2002; Sederberg et al., 2008), and the Tolman-
310 Eichenbaum machine in unifying spatial and nonspatial tasks (Whittington et al., 2020). A
311 common conceptual element of these models is the associative maps stored in memory and
312 employed to predict the future. Our observations are direct evidence that associative-pattern
313 learning from experience results in changes to the processing of such regularities. SL as a
314 common mechanism underlying the multifaceted hippocampal activity in navigation, cognitive
315 map, memory and learning is worth exploring in future work.

316 Moreover, the decoding results across mini blocks (Figs. 4B & C) suggest that feedback
317 from the environment shapes processing biases to better accommodate the true associative
318 statistics. As learning unfolded, the decoded transition matrices deviated from the initial patterns
319 and fluctuated across both S- and R-blocks. At the end, a triplet-like pattern was decoded in the

last S-blocks and but not the R-blocks (Fig. 4C). This decoding pattern depicts a learning process that does not progress in a linear fashion. A possible interpretation is that learners are particularly sensitive to the temporal statistics during the initial encountering of the stimuli (Jungé et al., 2017) and may "over-generalize" after this brief exposure (Kemp et al., 2007; Perfors et al., 2010; Wonnacott, 2011). The triplet-like decoded pattern in early R-blocks could reflect this strategy. The varying decoding efficacy in later blocks likely reflects the adjustment of learners' representations by repeated exposure to the same stimuli. The mini block design in this task can be a general experimental approach to further explore how learning evolves across stages.

Finally, our results suggest different roles of hippocampal subfields in SL and the resulting experience-driven changes to TP processing. The CA2/3 and CA4/DG showed the greatest task-average S-R difference in the BOLD autocorrelation, while the CA1 and parasubiculum showed sensitivity to the triplet TP pattern at the beginning and end of the learning session, respectively (Fig. 5C & D). These results intersect with the proposed functional dichotomy between the monosynaptic and trisynaptic pathways (O'Reilly & Norman, 2002; Turk-Browne, 2019). On one hand, the sensitivity of the CA2/3 and CA4/DG to the general structural difference between sequences fits their role in the trisynaptic pathway for pattern separation (Leutgeb et al., 2007). On the other hand, the CA1 is postulated to explicitly code for associative patterns in the monosynaptic pathway through SL (O'Reilly & Norman, 2002; Schapiro et al., 2017), which can account for its TP sensitivity in this study. Indeed, previous work has observed CA1 place-cell coding for environmental context based on experience (Plitt & Giocomo, 2021), supporting the history-dependent TP processing in our results. We postulate that the learned changes to TP processing in the CA1 is conveyed to the parasubiculum for output to downstream inferotemporal and visual regions. This output could alter the tuning for

343 objects in these regions through prediction-based processing adaptations to facilitate target
344 detection (Turk-Browne, 2019). The current study cannot test these hypotheses due to limited
345 temporal resolution of BOLD fMRI. Nonetheless, these results encourage further investigation
346 into the orchestrated dynamics among hippocampal subfields during SL.
347

348 **Online Methods**

349 *Participants.* Twenty-eight adults (mean age = 20.79 years, SD = 2.89 years, 7 males)
350 participated in this study. All participants were right-handed, native English speakers with no
351 history of neurological or psychiatric disorders. All participants gave written consent to perform
352 in accordance with the Institutional Review Board at the University of Delaware. All participants
353 were compensated for their participation. Due to incomplete MRI data resulting from acquisition
354 issues (N = 2), early termination of the study by the participants (N = 2), poor performance with
355 < 20% hit rate in at least one scanning session (N = 4), twenty participants in total (mean age =
356 20.16 years, SD = 0.81 years, 7 males) were included in the data analyses.

357 *Stimuli.* The stimuli contained two types of images: *Letters* (linguistic stimuli) were 12
358 images each containing an alien character holding up a sign with a capital letter (see
359 Supplemental Figure for all the stimuli). The alien character was held the same while the letter
360 differed among the 12 images. *Pictures* (nonlinguistic stimuli) were 12 different standalone alien
361 cartoons (Supplemental Figure). Both Letters and Pictures were presented in two types of mini
362 blocks. An *S-block* was composed of 48 stimuli, organized into 4 unique triplets each repeated 4
363 times. The presentation order was pseudo-randomized among triplets, so that the TP between
364 stimuli was 1 within-triplet and 1/4 between-triplet. No Letter triplet contained any words,
365 common acronyms, or initialisms. An *R-block* contained 12 unique stimuli each repeated 4 times.

366 In other words, an R-block was the randomization of all stimuli in an S-block of the same
367 stimulus type. The presentation order of individual stimuli was pseudo-randomized, and no
368 combination of three stimuli was presented more than once. Thus, the TP among stimuli was
369 uniformly 1/12. For both S- and R-blocks, each stimulus was presented at the center of the screen
370 for 0.8 s with an 0.2 s interval between stimuli. Each mini block lasted 48 s.

371 *Procedure.* Participants watched two sessions of stimuli presentation, organized into
372 mixed S- and R-blocks described above, while performing a target detection task. Session 1
373 contained 6 linguistic S-blocks, 6 nonlinguistic R-blocks and 6 12-second resting periods;
374 Session 2 contained 6 nonlinguistic S-blocks, 6 linguistic R-blocks and 6 12-second resting
375 periods. The order of S-blocks, R-blocks and resting periods was pseudo-randomized, and the
376 randomization was independent for each session and each participant. Adjustment was made to
377 ensure that no more than two blocks of the same type were adjacent to each other. Before each
378 block, a cue for the target was presented at the center of the screen for 2 s. For both linguistic and
379 nonlinguistic stimuli, the cue was a stimulus chosen from the last position of a triplet in an S-
380 block. The cue choice was randomized among participants. Throughout the task, in each
381 participant and for each type of stimuli, the same cue was used for all S- and R-blocks. Because
382 the order of stimuli and that of blocks were randomized independently for each participant, the
383 target location among stimuli followed no systematic pattern either within- or cross-participants.

384 *Structural and functional MRI data acquisition.* Data were acquired on a Siemens 3T
385 Magnetom Prisma scanner with a 64-channel phased array head coil at the Center for Brain and
386 Biomedical Imaging at The University of Delaware. Whole-head, high resolution structural
387 images, including a T1-weighted, magnetizations-prepared rapid gradient-echo (MPRAGE)
388 anatomical volume (TR=2500 ms, TE=2.9 ms, TI=1070 ms, flip angle=8.0°, voxel

389 resolution=1.0 mm isotropic, FOV=256 × 256, 176 sagittal slices) and a T2-weighted anatomical
390 volume (TR=3200 ms, TE=565 ms, flip angle=2.0°, voxel resolution=1.0 mm isotropic,
391 FOV=256 × 256, 32 sagittal slices) were collected. Functional images were acquired using
392 simultaneous multi-slice, T2*-weighted echo-planar imaging scans (TR=800 ms, TE=32 ms, flip
393 angle=61°, FOV=21 cm, matrix=64 × 64, acceleration factor=6). We acquired sixty adjacent
394 slices in an interleaved sequence with 2.5 mm slice thickness resulting in an in-plane resolution
395 of 2.5 × 2.5 × 2.5 mm^3.

396 *Image preprocessing.* MRI data were preprocessed using the analysis pipeline of
397 FreeSurfer (7.1.0) (Fischl, 2012) involving intensity normalization, registration to Talairach
398 space, skull stripping, segmentation of white matter, tessellation of the white matter boundary,
399 and automatic correction of topological defects (Fischl & Dale, 2000). BOLD preprocessing used
400 a combination of SPM12 (Penny et al., 2011), FSL (5.0.9) (Smith et al., 2001), AFNI (Cox &
401 Hyde, 1997) and FreeSurfer scripts implemented in Nipype 1.1.9 (Gorgolewski et al., 2011).
402 Preprocessing steps involved dropping the 4 initial frames, despiking (AFNI), slice time
403 correction (SPM SliceTiming), motion correction (SPM Realign), removal of linear and
404 quadratic trends (FSL TSNR), and artifact detection based on movement or deviation in intensity
405 (Nipype ArtifactDetect). Because of our subsequent autocorrelation and HMM analyses, no
406 smoothing was applied to avoid distortion in the BOLD time series. Lastly, each subject's T1
407 volume was co-registered to the functional space using boundary-based registration with 9
408 degrees of freedom (FreeSurfer bbregister).

409 *Segmentation of the hippocampal subregions.* Hippocampal subregions were segmented
410 from the T1-weighted images using an automated algorithm provided by FreeSurfer 7.1.0. This
411 algorithm utilizes a probabilistic atlas built upon ultra-high resolution (0.13mm isotropic, on

average), ex vivo MRI data from 15 cases (Iglesias et al., 2015). The segmentation yielded the following subregions: parasubiculum, presubiculum, subiculum, CA1, CA3, CA4, granule cell layer of dentate gyrus (DG), hippocampus-amygdala-transition-area (HATA), fimbria, molecular layer, hippocampal fissure, and tail. The HATA and following subregions, as well as the parasubiculum, are fine-grained structures that occupy only a few voxels. Thus, we grouped parasubiculum and presubiculum into one subregion "parasubiculum", and CA4, DG, HATA and following subregions into one "CA4/DG".

BOLD autocorrelation. BOLD images were coregistered to the FreeSurfer preprocessed anatomical images in each subject's native space. BOLD time series were extracted from hippocampal voxels using the segmentation masks. Each data point is the closest after stimulus onset. For each voxel, lag-1 autocorrelation r was calculated following the formula:

$$r = \frac{[X(t) - \bar{X}] \cdot [X(t) - \bar{X}]}{\|X(t) - \bar{X}\| \cdot \|X(t-1) - \bar{X}\|}, \qquad (1)$$

where $X(t) = [x_1, x_2, x_3, \ldots, x_T]$ is a vector of BOLD activity at time points 1 to T, $X(t\text{-}1)$ denotes the BOLD activity shifted by one time step, $\bar{X}$ is the temporal mean of $X(t)$, $\|\|$ denotes the vector norm, and dot denotes inner product. Each mini block contained 48 stimuli, and thus, 48 time points in the BOLD time series, and 47 time points after balancing for the lag. For cross-subject comparison in Fig. 1, the BOLD autocorrelation was resampled to the MNI152 T1 1mm template in FSL using FreeSurfer's mri_robust_register algorithm. For all other analyses the BOLD autocorrelation was extracted from each subject's native space.

Linear mixed models and ANOVA. Within-subject treatment effects were analyzed with linear mixed models (Bates et al., 2015). Repeated-measures ANOVA were applied to the model output for assessing effect size and significance, with type III sums of squares and the

Satterthwaite approximation for degrees of freedom. Implementation of these analyses used the lme4 package in R (4.0.5) with the RStudio interface (1.4.1106; Rstudio Inc.).

Hidden Markov modeling. The intuition behind this HMM came from the hypothesis that hippocampal neurons should be sensitive to the TP between stimuli regardless of their sensory features. Thus, the hidden states of this HMM contained 3 variables to represent the triplet sequence:

$$Q = \{q_1, q_2, q_3\}, \qquad (2)$$

where q_1, q_2, q_3 denote the first (*start*), second (*mid*) and third (*end*) positions in a triplet. A 48-stimulus triplet sequence would correspond to 16 repetitions of the $q_1q_2q_3$ set. The probability of transition between the hidden states was defined by the matrix

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}. \qquad (3)$$

Each element on the i-th row and j-th column represents the probability of transition from q_i to q_j. By construction, the transition for a triplet sequence was only possible from *start* to *mid*, *mid* to *end*, and *end* to *start*. Thus, A had a probability of 1 in the cells corresponding to these transitions and 0 elsewhere. Because a structured sequence always begins with *start*, the initial probability of Q was set to $\pi = [1, 0, 0]$. Together, Q and A define the triplet sequence that drove the SL-related BOLD activity. The goal of the HMM analysis was to test whether A can be decoded from the observed BOLD activity.

In the HMM framework, observations are linked to hidden states via an emission probability function. For the analysis in this study, the BOLD signal was averaged across hippocampal voxels (within subfields for the subfield analysis) and treated as a univariate time series, $O(t)$. The relationship between the hidden states and $O(t)$ was determined by the emission

probability function $b(o_t|q_j)$, which represented the probability of observing activity o at time t given state j. Critically, $b(o_t|q_j)$ was hypothesized to follow a conditional Gaussian distribution:

$$b(o_t|q_j) \sim N(m_j + \rho_{ij}(o_{t-1} - m_i), (1 - \rho_{ij}^2)\sigma^2), \quad (4)$$

where m_i and m_j denote the mean BOLD activity corresponding to hidden states i and j, respectively, ρ_{ij} the correlation of the BOLD activity corresponding to hidden states i and j, and σ the standard deviation that was assumed to be constant in the BOLD time series. The right-hand side of Eq. 4 is the conditional distribution of o_t (corresponding to state j) given o_{t-1} (corresponding to state i), assuming a bivariate Gaussian joint distribution. Note that o_t and o_{t-1} denote measurements at adjacent time points, and therefore their correlation ρ_{ij} is the lag-1 autocorrelation of the BOLD activity. Eq. 4 states that given hidden state j, the probability of observing o_t was a function of state-dependent variables m_i, m_j and ρ_{ij}. For triplet sequences, because the transitions among states were deterministic, the i-j pairs were unique, i.e., there was a unique m_i for each m_j. Thus, $b(o_t|q_j)$ was unique for each state j at time t.

The presence of autocorrelation ρ_{ij} in $b(o_t|q_j)$ set apart our conditional-Gaussian HMM from previous univariate-Gaussian HMMs that had been applied to task fMRI studies (Baldassano et al., 2017; Duan & Man, 2012). In a univariate-Gaussian framework, the setup for hidden states Q, transition matrix A and initial states π would be the same as above, but the emission probability would depend only on the mean and variance of the BOLD signal at time t:

$$b_{univ}(o_t|q_j) \sim N(m_j, \sigma^2). \quad (5)$$

In Eq. 5, the only state-dependent parameter was the mean BOLD activity m_j. The univariate-Gaussian HMM assumes that different hidden states resulted in different mean levels of BOLD activity. In contrast, the conditional-Gaussian HMM relaxes this assumption by adding influences of autocorrelation ρ_{ij} to $b(o_t|q_j)$. In Eq. 4, the hidden states could be linked to o_t via

either the mean values or ρ_{ij} (or both). For example, if all three states corresponded to the same mean BOLD activity, then ρ_{ij} would become the only state-dependent parameter to affect the emission probability, and the hidden state information would be entirely encoded in the autocorrelation. Alternatively, if SL did not affect the autocorrelation, i.e. $\rho_{ij} = 0$ for all state pairs, then Eq. 4 would become Eq. 5 and the hidden state information would be entirely encoded in the mean values. Thus, the conditional-Gaussian HMM allowed us to test whether the BOLD autocorrelation contained any information of the hidden triplet sequence.

The test was achieved through a decoding procedure, in which the hidden-state transition matrix A was treated as an unknown parameter to be estimated, or decoded, from the observed data. By comparing the decoded matrix with the ground truth in Eq. 3, one could assess the model efficacy in capturing the relationship between the hidden states and the BOLD data. If the conditional-Gaussian HMM yielded more accurate results than the univariate-Gaussian HMM, it would be evidence that the relationship specified by Eq. 4 was more plausible than that by Eq. 5, i.e., the BOLD autocorrelation contributed to encoding the hidden state information. The accuracy of the decoded matrix would reflect the accuracy of the hypothesized encoding scheme in Eq. 4. For the decoding procedure, the definition of our conditional-Gaussian HMM fulfills conventional HMM assumptions and its parameters can be estimated via the standard Baum-Welch algorithm (Jurafsky & Martin, 2009). To initiate this algorithm, an uninformative prior of A was set up such that each element in the matrix was 1/3. The prior of $b(\cdot)$ was set up using parameters calculated from the empirical BOLD data. For each mini block, the observation $O(t)$ contained 48 BOLD data points corresponding to the 48 stimuli:

$$O_t = \{o_1, o_2, o_3, \ldots, o_{48}\}, \quad (6)$$

The initial values of parameters in Eq. 4 were calculated from O_t:

$$\hat{m}_i = \frac{1}{16} \sum_{\forall t: q_t = q_i} o_t \quad , \qquad (7)$$

$$\hat{\sigma} = \sqrt{\frac{1}{47} \sum_{t=1}^{48} (o_t - \hat{m})^2} \quad , \qquad (8)$$

where $\hat{m}_i$ denotes the mean of o_t across time points corresponding to state i, and $\hat{m}$ the mean across all time points. The initial value of ρ_{ij} was calculated separately for each i-j pair:

$$\hat{\rho}_{ij} = \frac{\sum_t (o_{t-1,i} - \hat{m}_i)(o_{t,j} - \hat{m}_j)}{\sqrt{\sum_t (o_{t-1,i} - \hat{m}_i)^2 \sum_t (o_{t,j} - \hat{m}_j)^2}} \quad , \qquad (9)$$

in which the summation was over all time points associated with state pair i-j. With the initial setup of A and $b(\cdot)$, the Baum-Welch algorithm will re-estimate A and $b(\cdot)$ recursively through an expectation-maximization procedure until the results converge. To compare the BOLD-estimated transition matrix with the ground-truth represented by Eq. 3, a difference score was computed in the form of a normalized mean squared error (n-MSE):

$$\text{n-MSE} = \sum_{i,j} (a_{ij} - \hat{a}_{ij})^2 / \sum_{i,j} a_{ij}^2 \quad , \qquad (10)$$

where a_{ij} denotes elements in the ground-truth matrix (Eq. 3), and $\hat{a}_{ij}$ elements in the estimated transition matrix.

Simulation. Stochastic processes were generated from the HMMs for testing decoding efficacy. HMMs are generative models, i.e. given π, A, and $b(\cdot)$, samples of o_t can be drawn from the conditional-Gaussian emission probability distribution up to an arbitrary time point T. To match the task setup, we let $T = 48$ for all simulated time series. To test the decoding efficacy

under different parameter settings, two quantities were systematically varied: the sensitivity index d' and autocorrelation ρ_{ij}. Sensitivity index is a function of m_i, m_j, and σ:

$$d' = (|m_1 - m_2| + |m_2 - m_3| + |m_3 - m_1|)/3\sigma. \qquad (11)$$

It measures the joint effect of m_i and m_j, i.e., how well the observations corresponding to the three states can be separated by their mean values. For each parameter set, 100 time series were generated and a Baum-Welch decoding algorithm was applied to each sample. The accuracy was measured by n-MSE averaged across the 100 samples.

Acknowledgement

This work was supported in part by National Institute on Deafness and Other Communication Disorders (NIDCD) Grant R21DC010576 to Z.Q. The data collection was conducted at the University of Delaware. The main analysis was conducted on the high-performance computing platform Big Red 3 and Carbonate, supported in part by Lilly Endowment, Inc. for the Indiana University Pervasive Technology Institute. We thank the staff at the Center for Biological and Brain Imaging at the University of Delaware for their critical support for our neuroimaging data collection. We thank Violet Kozloff and An Nguyen for their assistance in stimulus construction and programming for the experiment, Julie Schneider, Jennifer Legault, and Yi-Lun Weng for their contribution in data collection and organization, Anqi Hu for her assistance in behavioral data processing, and Marc Howard, Samantha Wood, Zoran Tiganj, Ho Ming Chow, Diane Chugani and Frances Sayako Earle for helpful comments and discussions on the manuscript.

References

Baldassano, C., Chen, J., Zadbood, A., Pillow, J. W., Hasson, U., & Norman, K. A. (2017).

Discovering Event Structure in Continuous Narrative Perception and Memory. *Neuron*, *95*(3), 709–721.e5.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Usinglme4. In *Journal of Statistical Software* (Vol. 67, Issue 1). https://doi.org/10.18637/jss.v067.i01

Batterink, L. J. (2017). Rapid Statistical Learning Supporting Word Extraction From Continuous Speech. *Psychological Science*, *28*(7), 921–928.

Batterink, L. J., & Paller, K. A. (2017). Online neural monitoring of statistical learning. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *90*, 31–45.

Bischoff-Grethe, A., Proper, S. M., Mao, H., Daniels, K. A., & Berns, G. S. (2000). Conscious and Unconscious Processing of Nonverbal Predictability in Wernicke's Area. In *The Journal of Neuroscience* (Vol. 20, Issue 5, pp. 1975–1981). https://doi.org/10.1523/jneurosci.20-05-01975.2000

Burgess, N., Maguire, E. A., & O'Keefe, J. (2002). The Human Hippocampus and Spatial and Episodic Memory. In *Neuron* (Vol. 35, Issue 4, pp. 625–641). https://doi.org/10.1016/s0896-6273(02)00830-9

Cohen, N. J., & Eichenbaum, H. (1995). *Memory, Amnesia, and the Hippocampal System*. MIT Press.

Conway, C. M. (2020). How does the brain learn environmental structure? Ten core principles for understanding the neurocognitive mechanisms of statistical learning. *Neuroscience and Biobehavioral Reviews*, *112*, 279–299.

Covington, N. V., Brown-Schmidt, S., & Duff, M. C. (2018). The Necessity of the Hippocampus for Statistical Learning. *Journal of Cognitive Neuroscience*, *30*(5), 680–697.

Cox, R. W., & Hyde, J. S. (1997). Software tools for analysis and visualization of fMRI data. *NMR in Biomedicine*, *10*(4-5), 171–178.

Dayan, P. (1993). Improving Generalization for Temporal Difference Learning: The Successor Representation. *Neural Computation*, *5*(4), 613–624.

Duan, R., & Man, H. (2012). Activation Detection on fMRI Time Series Using Hidden Markov Model. In *Advances in Artificial Neural Systems* (Vol. 2012, pp. 1–12). https://doi.org/10.1155/2012/190359

Ekstrom, A. D., & Ranganath, C. (2018). Space, time, and episodic memory: The hippocampus is all over the cognitive map. *Hippocampus*, *28*(9), 680–687.

Fischl, B. (2012). FreeSurfer. *NeuroImage*, *62*(2), 774–781.

Fischl, B., & Dale, A. M. (2000). Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(20), 11050–11055.

Fiser, J., & Aslin, R. N. (2002). Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *28*(3), 458–467.

Frost, R., Armstrong, B. C., Siegelman, N., & Christiansen, M. E. (2015). Domain generality versus modality specificity: the paradox of statistical learning. In *Trends in Cognitive Sciences*, *19*(3), 117–125.

Furl, N., Kumar, S., Alter, K., Durrant, S., Shawe-Taylor, J., & Griffiths, T. D. (2011). Neural prediction of higher-order auditory sequence statistics. In *NeuroImage* (Vol. 54, Issue 3, pp. 2267–2277). https://doi.org/10.1016/j.neuroimage.2010.10.038

Gorgolewski, K., Burns, C. D., Madison, C., Clark, D., Halchenko, Y. O., Waskom, M. L., &

Ghosh, S. S. (2011). Nipype: A Flexible, Lightweight and Extensible Neuroimaging Data Processing Framework in Python. In *Frontiers in Neuroinformatics* (Vol. 5). https://doi.org/10.3389/fninf.2011.00013

Howard, M. W., & Kahana, M. J. (2002). A Distributed Representation of Temporal Context. *Journal of Mathematical Psychology*, *46*(3), 269–299.

Huang, Z., Liu, X., Mashour, G. A., & Hudetz, A. G. (2018). Timescales of Intrinsic BOLD Signal Dynamics and Functional Connectivity in Pharmacologic and Neuropathologic States of Unconsciousness. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *38*(9), 2304–2317.

Iglesias, J. E., Augustinack, J. C., Nguyen, K., Player, C. M., Player, A., Wright, M., Roy, N., Frosch, M. P., McKee, A. C., Wald, L. L., Fischl, B., Van Leemput, K., & Alzheimer's Disease Neuroimaging Initiative. (2015). A computational atlas of the hippocampal formation using ex vivo, ultra-high resolution MRI: Application to adaptive segmentation of in vivo MRI. *NeuroImage*, *115*, 117–137.

Isbilen, E. S., McCauley, S. M., Kidd, E., & Christiansen, M. H. (2020). Statistically Induced Chunking Recall: A Memory-Based Approach to Statistical Learning. *Cognitive Science*, *44*(7), e12848.

Jungé, J. A., Scholl, B. J., & Chun, M. M. (2007). How is spatial context learning integrated over signal versus noise? A primacy effect in contextual cueing. *Visual Cognition*, *15*(1), 1–11.

Jurafsky, D., & Martin, J. H. (2009). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Prentice Hall.

Karuza, E. A., Newport, E. L., Aslin, R. N., Starling, S. J., Tivarus, M. E., & Bavelier, D. (2013).

The neural correlates of statistical learning in a word segmentation task: An fMRI study. *Brain and Language*, *127*(1), 46–54.

Kemp, C., Perfors, A., & Tenenbaum, J. B. (2007). Learning overhypotheses with hierarchical Bayesian models. *Developmental Science*, *10*(3), 307–321.

Leutgeb, J. K., Leutgeb, S., Moser, M.-B., & Moser, E. I. (2007). Pattern separation in the dentate gyrus and CA3 of the hippocampus. *Science*, *315*(5814), 961–966.

McNealy, K., Mazziotta, J. C., & Dapretto, M. (2006). Cracking the language code: neural mechanisms underlying speech parsing. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *26*(29), 7629–7639.

O'Keefe, J., Nadel, L., & Regents Professor of Psychology Lynn Nadel. (1978). *The Hippocampus as a Cognitive Map*. Oxford University Press, USA.

O'Reilly, R. C., & Norman, K. A. (2002). Hippocampal and neocortical contributions to memory: advances in the complementary learning systems framework. *Trends in Cognitive Sciences*, *6*(12), 505–510.

Penny, W. D., Friston, K. J., Ashburner, J. T., Kiebel, S. J., & Nichols, T. E. (2011). *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. Elsevier.

Perfors, A., Tenenbaum, J. B., & Wonnacott, E. (2010). Variability, negative evidence, and the acquisition of verb argument constructions. *Journal of Child Language*, *37*(3), 607–642.

Plitt, M. H., & Giocomo, L. M. (2021). Experience-dependent contextual codes in the hippocampus. *Nature Neuroscience*, *24*(5), 705–714.

Reber, P. J. (2013). The neural basis of implicit learning and memory: a review of neuropsychological and neuroimaging research. *Neuropsychologia*, *51*(10), 2026–2042.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical Learning by 8-Month-Old

Infants. In *Science* (Vol. 274, Issue 5294, pp. 1926–1928). https://doi.org/10.1126/science.274.5294.1926

Schapiro, A. C., Gregory, E., Landau, B., McCloskey, M., & Turk-Browne, N. B. (2014). The necessity of the medial temporal lobe for statistical learning. *Journal of Cognitive Neuroscience*, *26*(8), 1736–1747.

Schapiro, A. C., Kustner, L. V., & Turk-Browne, N. B. (2012). Shaping of object representations in the human medial temporal lobe based on temporal regularities. *Current Biology: CB*, *22*(17), 1622–1627.

Schapiro, A. C., Rogers, T. T., Cordova, N. I., Turk-Browne, N. B., & Botvinick, M. M. (2013). Neural representations of events arise from temporal community structure. *Nature Neuroscience*, *16*(4), 486–492.

Schapiro, A. C., Turk-Browne, N. B., Botvinick, M. M., & Norman, K. A. (2017). Complementary learning systems within the hippocampus: a neural network modelling approach to reconciling episodic memory with statistical learning. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *372*(1711). https://doi.org/10.1098/rstb.2016.0049

Schiller, D., Eichenbaum, H., Buffalo, E. A., Davachi, L., Foster, D. J., Leutgeb, S., & Ranganath, C. (2015). Memory and Space: Towards an Understanding of the Cognitive Map. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *35*(41), 13904–13911.

Schneider, J. M., Hu, A., Legault, J., & Qi, Z. (2020). Measuring Statistical Learning Across Modalities and Domains in School-Aged Children Via an Online Platform and Neuroimaging Techniques. *Journal of Visualized Experiments: JoVE*, *160*.

https://doi.org/10.3791/61474

Sederberg, P. B., Howard, M. W., & Kahana, M. J. (2008). A context-based theory of recency and contiguity in free recall. *Psychological Review*, *115*(4), 893–912.

Sherman, B. E., & Turk-Browne, N. B. (2020). Statistical prediction of the future impairs episodic encoding of the present. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(37), 22760–22770.

Smith, S., Bannister, P. R., Beckmann, C., Brady, M., Clare, S., Flitney, D., Hansen, P., Jenkinson, M., Leibovici, D., Ripley, B., Woolrich, M., & Zhang, Y. (2001). FSL: New tools for functional and structural brain image analysis. *NeuroImage*, *13*(6), 249.

Stachenfeld, K. L., Botvinick, M. M., & Gershman, S. J. (2017). The hippocampus as a predictive map. *Nature Neuroscience*, *20*(11), 1643–1653.

Thothathiri, M., & Rattinger, M. (2015). Controlled processing during sequencing. In *Frontiers in Human Neuroscience* (Vol. 9). https://doi.org/10.3389/fnhum.2015.00599

Turk-Browne, N. B. (2019). The hippocampus as a visual area organized by space and time: A spatiotemporal similarity hypothesis. *Vision Research*, *165*, 123–130.

Turk-Browne, N. B., Johnson, M. K., Chun, M. M., & Scholl, B. J. (2010). Neural evidence of statistical learning: Incidental detection and anticipation of regularities. In *Journal of Vision* (Vol. 8, Issue 6, pp. 695–695). https://doi.org/10.1167/8.6.695

Turk-Browne, N. B., Jungé, J., & Scholl, B. J. (2005). The automaticity of visual statistical learning. *Journal of Experimental Psychology. General*, *134*(4), 552–564.

Turk-Browne, N. B., Scholl, B. J., Chun, M. M., & Johnson, M. K. (2009). Neural evidence of statistical learning: efficient detection of visual regularities without awareness. *Journal of Cognitive Neuroscience*, *21*(10), 1934–1945.

Whittington, J. C. R., Muller, T. H., Mark, S., Chen, G., Barry, C., Burgess, N., & Behrens, T. E. J. (2020). *The Tolman-Eichenbaum Machine: Unifying space and relational memory through generalisation in the hippocampal formation*. *Cell*, *183*(5), 1249–1263.

Wonnacott, E. (2011). Balancing generalization and lexical conservatism: An artificial language study with child learners. *Journal of Memory and Language*, *65*(1), 1–14.

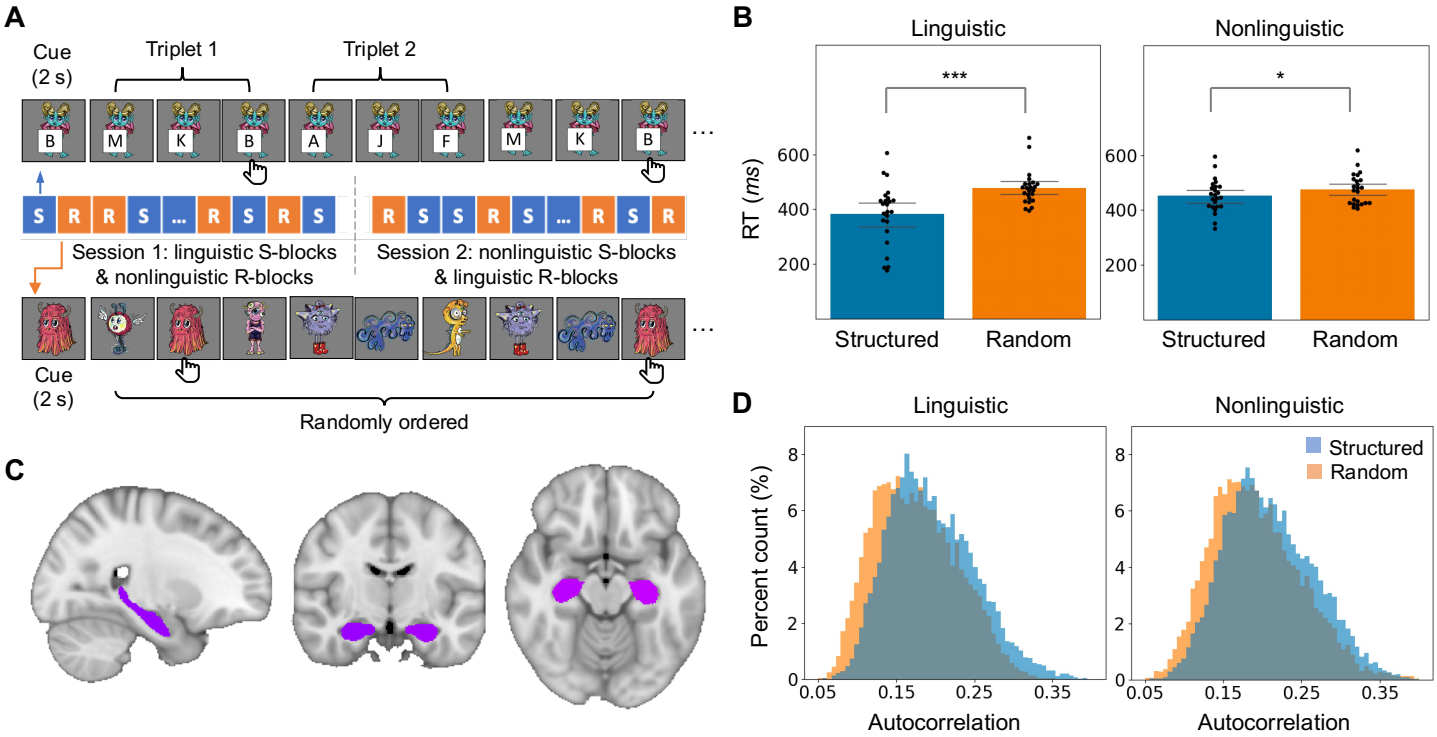


Figure 1. Repeated exposure to structured stimuli corresponded with reduced RT for target detection and increased lag-1 autocorrelation in the hippocampal activity. (A) Schematic of the task design. Subjects viewed 2 sessions of stimuli presentation, each composed of 12 randomly intermixed mini blocks (middle row). Session 1 contained 6 S-blocks (blue) of linguistic stimuli and 6 R-blocks (orange) of nonlinguistic stimuli, and in Session 2 the S/R block type was switched between linguistic and nonlinguistic stimuli. Each mini block started with a cue for the target, followed by 48 stimuli presented sequentially. Each stimulus was presented at the center of the screen for 0.8 s, followed by a 0.2 s inter-stimulus interval. Subjects responded to the target with a button press. Top row: an example S-block, in which linguistic stimuli were organized into triplets. Bottom row: an example R-block, in which nonlinguistic stimuli were randomly ordered. (B) RT of the structured and random sequences for each type of stimuli, shown as median ± s.e. across subjects. Black dots on the swarm plot represent the cross-trial median of individual subjects. Asterisks mark the significance level of the Wilcoxon test, ***: $p < 0.001$, *: $p < 0.05$. (C) The hippocampus mask on the MNI152 T1 brain template from FSL. (D) Histograms of BOLD autocorrelation across hippocampal voxels in the MNI152 template. The BOLD autocorrelation from each subject's native MRI space was transformed to the MNI152 template. The autocorrelation value of each voxel in the template was averaged across subjects.

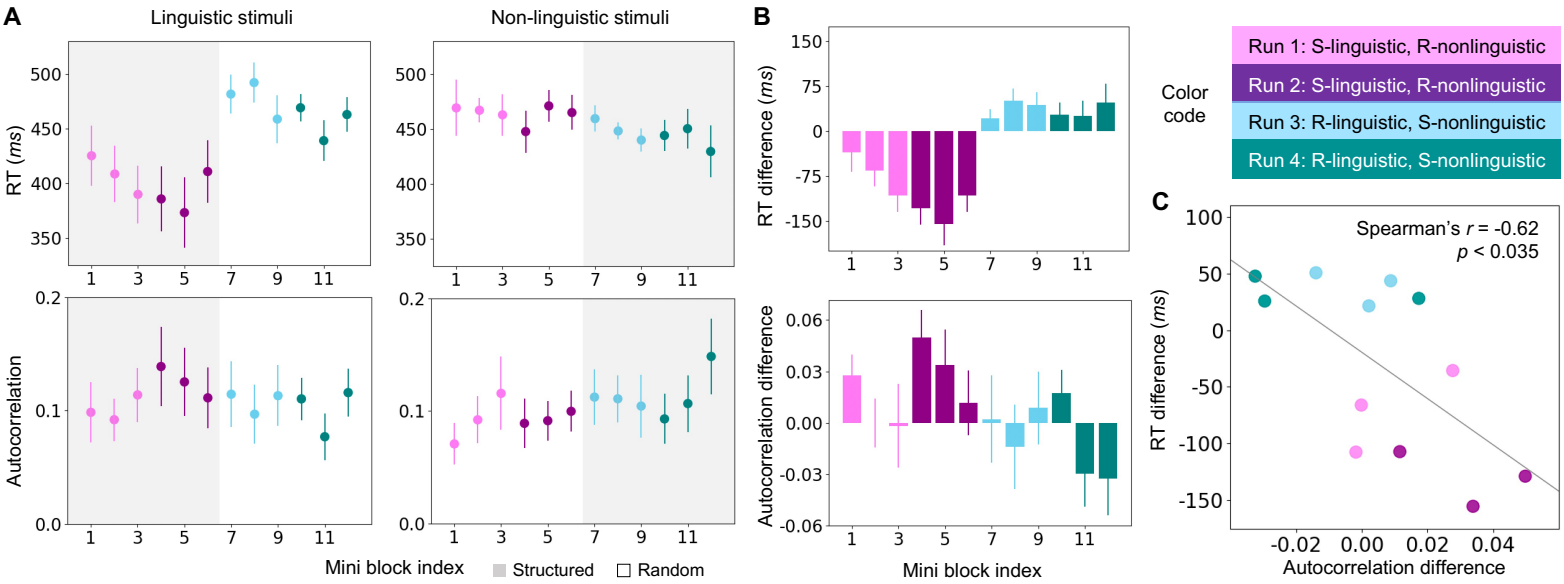


Figure 2. Behavioral and hippocampal SL effects were correlated across mini blocks. (A) RT (median ± s.e. across subjects) and BOLD autocorrelation (mean ± s.e. across subjects) varied across mini blocks. The blocks were numbered by the order of presentation throughout the entire task. Background color separates the two sessions, with gray indicating the session in which the stimuli were structured. (B) The S-R difference of RT (median ± s.e. across subjects) and BOLD autocorrelation (mean ± s.e. across subjects) showed a sign-flip across sessions. The mini blocks were numbered in the same manner as in A. (C) The S-R difference of RT and that of BOLD autocorrelation were linearly correlated across mini blocks.

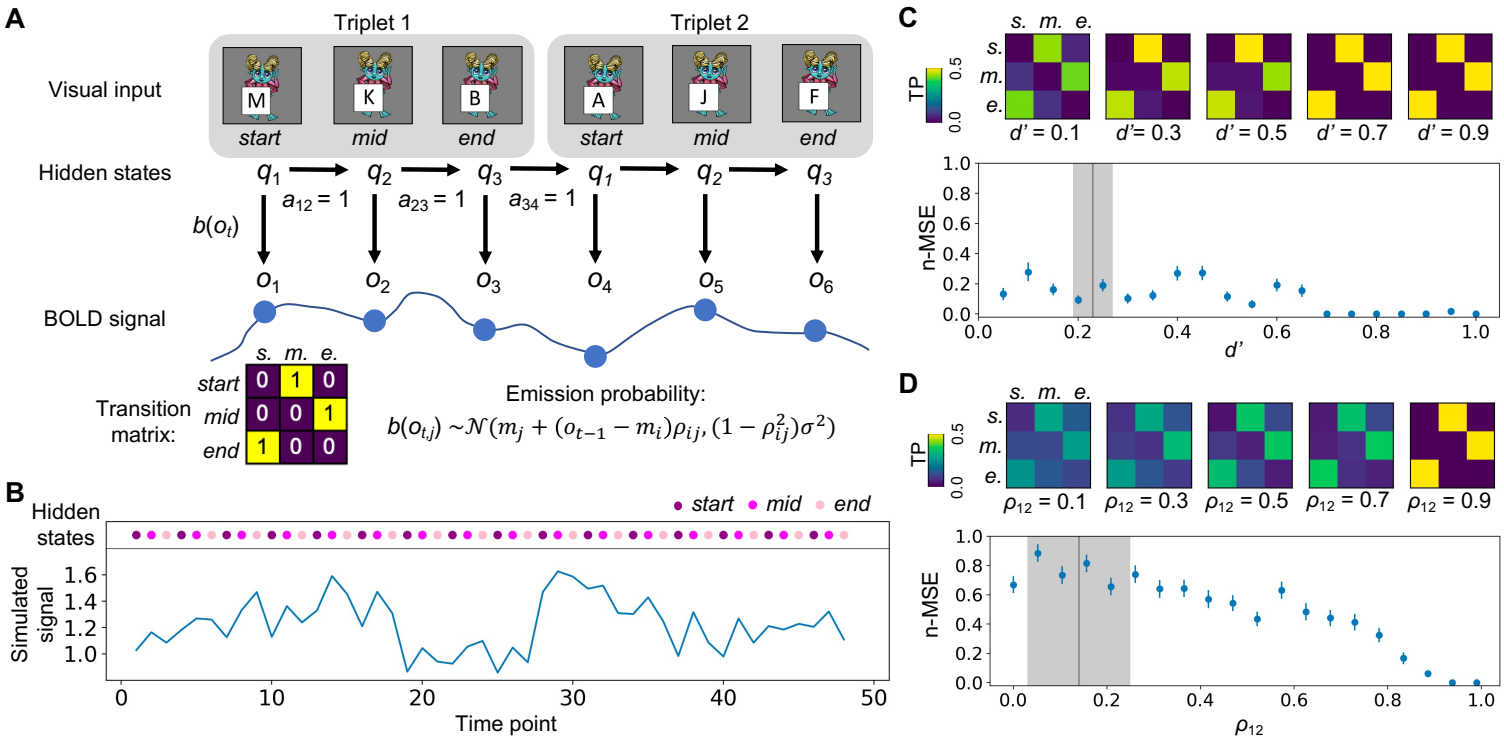


Figure 3. Conditional-Gaussian HMM decoding with simulated data demonstrated the model efficiency in capturing the triplet transition pattern. (A) Schematic of the conditional-Gaussian HMM structure. Three hidden states *start, mid, end* are defined according to the position of each stimulus in a triplet, regardless of their specific visual features. The TP among states is defined by the transition matrix (abbreviations: s. for start, m. for mid, e. for end). Each hidden state is linked to the observed BOLD signal by the emission probability, which follows a conditional-Gaussian distribution. (B) A simulated time series from the HMM, with parameters $m_1 = 0.4$, $m_2 = 0.5$, $m_3 = 0.6$, $\sigma = 0.1$, $\rho_{12} = \rho_{23} = 0.8$, $\rho_{31} = 0.2$. (C) Decoded transition matrices (top) and decoding accuracy (bottom) with varying d' and $\rho_{12} = \rho_{23} = 0.8$, $\rho_{31} = 0.2$, averaged across 100 runs. Error bars represent s.e. across runs. Shaded area marks the mean and s.d. of d' in the veridical BOLD data. (D) Decoded transition matrices (top) and decoding accuracy (bottom) with varying $\boldsymbol{\rho}$ and $d' = 0.23$, while keeping the relationship $\rho_{12} = \rho_{23} = 4\rho_{31}$. The results were averaged across runs, with error bars representing s.e. across runs. Shaded area marks the mean and s.d. of $\boldsymbol{\rho}$ in the veridical BOLD data.

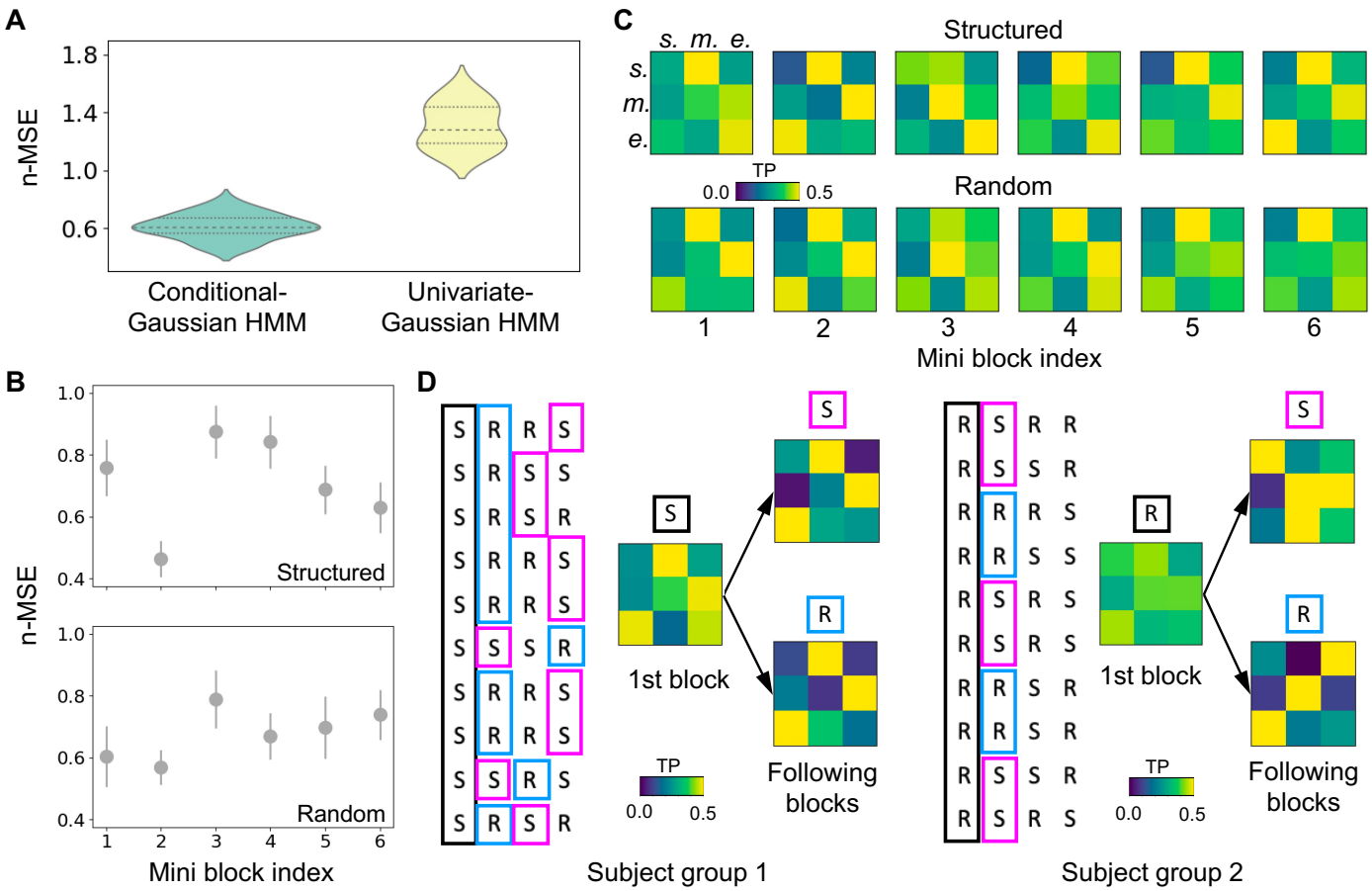


Figure 4. HMM decoding with veridical BOLD data showed varying accuracy across the mini blocks. (A) Decoding accuracy was significantly higher from the conditional-Gaussian than the univariate-Gaussian HMM. The symmetric contours represent the subject-wise histograms of n-MSE. Dashed lines mark the first, second and third quantiles of the distribution. (B) Decoding accuracy (mean ± s.e. across subjects) significantly varied across mini blocks. The n-MSE scores were averaged across sessions. (C) Decoded transition matrices for each mini block, averaged across sessions. Effective decoding was present in the early (first and/or second) S- and R-blocks, as well as the last S-block. (D) Decoding outcomes showed a mere-exposure effect. The letter arrays list the structure/randomness of the first 4 mini blocks (columns) in each subject (rows). The subjects were divided into two groups according to the stimulus structure in the first mini block (black boxes). For Group 1, decoded matrices are shown for the first S-block, and the first S- (magenta boxes) and R-blocks (blue boxes) that followed the first S-block, averaged across subjects. For Group 2, decoded matrices are shown for the first R-block, and the S- and R-blocks that immediately followed the first R-block, averaged across subjects. The selection strategy for Group 2 was based on the concern that any S-block between the first and second R-blocks might affect the hippocampal response to the second R-block, according to the mere-exposure assumption. Therefore, only the S- and R-blocks that immediately followed the first R-block were chosen for decoding comparison. Group 1 showed effective decoding in all selected blocks while Group 2 showed none.

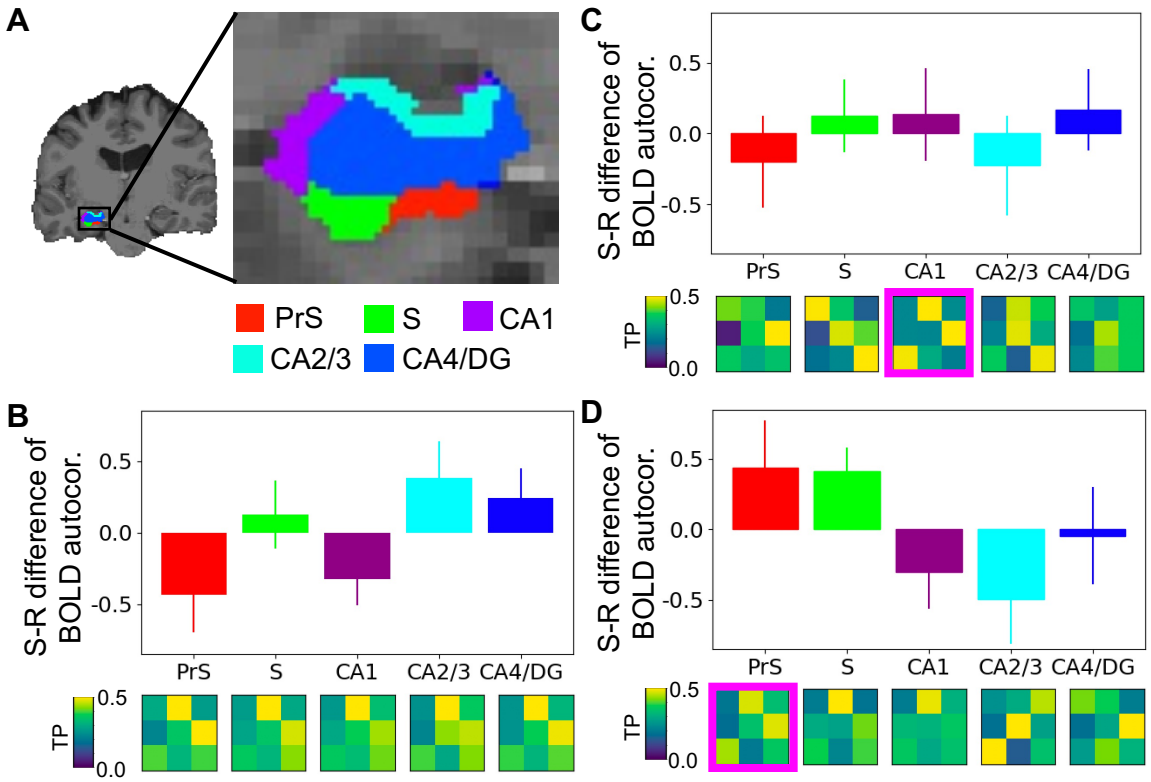
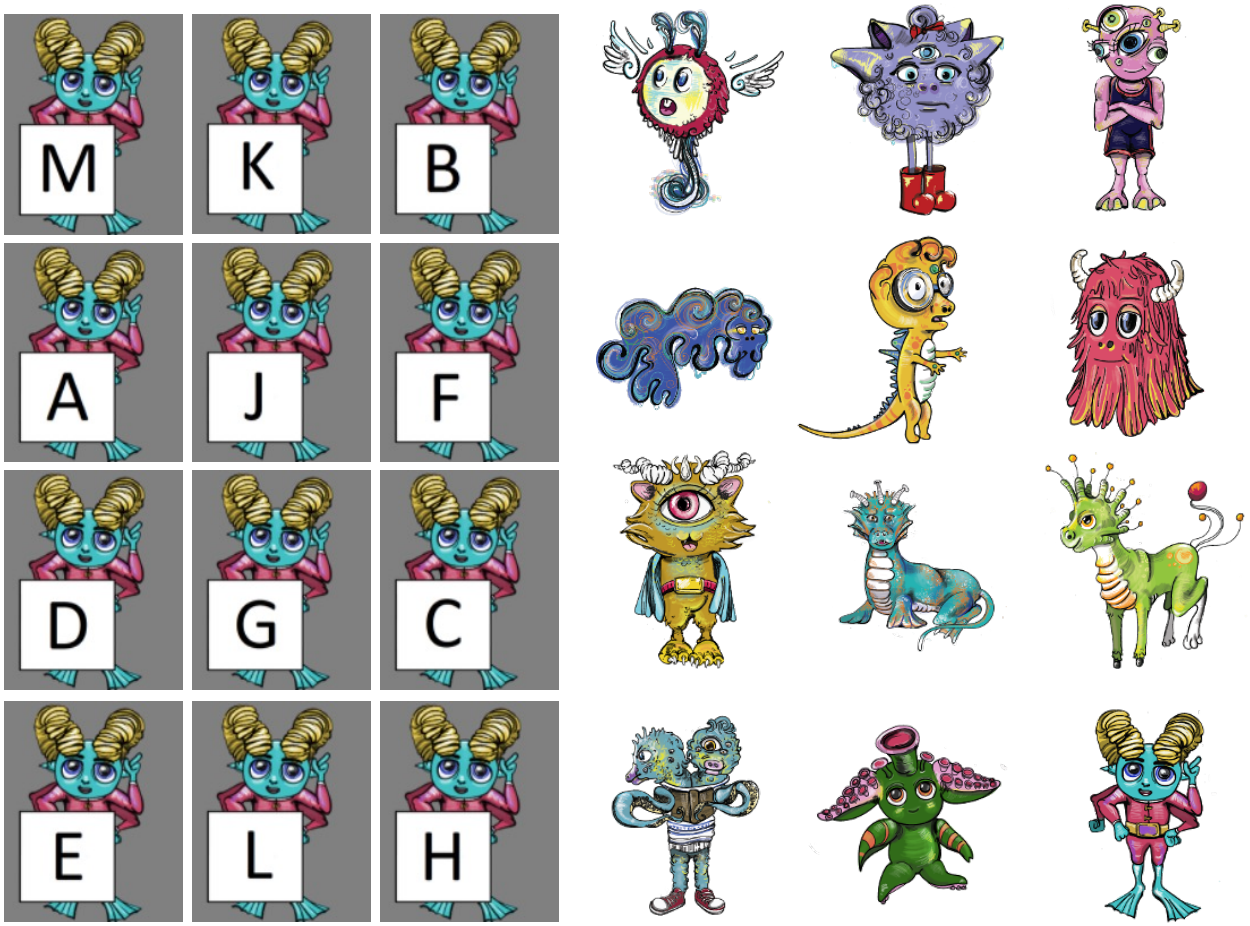


Figure 5. SL effects differed across hippocampal subfields. (A) Segmentation of hippocampal subfields shown in a subject's native space. Abbreviations: PrS=parasubiculum, S=subiculum. (B-D) S-R difference of BOLD autocorrelation (*z*-normalized, i.e., mean-subtracted and divided by s.d. across subfields) and decoded transition matrix for each subfield, using data averaged across all mini blocks (B) and from the first (C) and last (D) mini blocks. Colors in the bar charts correspond to the color scheme in A. Magenta boxes highlight effective decoding results.



Supplemental Figure. All the linguistic (left) and nonlinguistic (right) stimuli used in the task. For each type of stimuli, the three items on each row represent a triplet used in S-blocks.