Brief Communication

Visual object recognition is facilitated by temporal community structure

Ehsan Kakaei,[1,2,3] Stepan Aleshin,[2,3] and Jochen Braun[2,3]

[1]European Structural and Investment Funds Graduate School on Analysis, Imaging, and Modelling of Neuronal and Inflammatory Processes , Otto-von-Guericke University, 39120 Magdeburg, Germany; [2]Institute of Biology, Otto-von-Guericke University, 39120 Magdeburg, Germany; [3]Center for Behavioral Brain Sciences, Otto-von-Guericke University, 39120 Magdeburg, Germany

Humans and others primates are highly attuned to temporal consistencies and regularities in their sensory environment and learn to predict such statistical structure. Moreover, in several instances, the presence of temporal structure has been found to facilitate procedural learning and to improve task performance. Here we extend these findings to visual object recognition and to presentation sequences in which mutually predictive objects form distinct clusters or "communities." Our results show that temporal community structure accelerates recognition learning and affects the order in which objects are learned ("onset of familiarity").

[Supplemental material is available for this article.]

Our understanding of the world is grounded in sensory experience. Typically, this experience consists of contiguous streams of sensations that are richly structured in both time and space (Schapiro and Turk-Browne 2015). Such statistical structure may involve simple correlations of pairs of sensory events or, more generally, clusters of correlations between mutually predictive events forming a "temporal community" (Schapiro et al. 2013). Both humans and other primates (Miyashita 1988) can learn to predict such statistical regularities in space and time (Fiser and Aslin 2001, 2002). Moreover, statistical structure can be exploited explicitly or implicitly to enhance task performance. For example, predictable presentation order can facilitate motor learning (Kahn et al. 2018), language learning (Saffran et al. 1996), visual search (Chun and Jiang 1998; Jiang and Wagner 2004; Sisk et al. 2019), and conditional associative learning (Hamid et al. 2010).

In general, implicit (unsupervised) learning of temporal structure is thought to provide a biological basis for important cognitive functions, including the formation of episodic memories, learning of task-sets, model-based planning, and structural learning (e.g., Kemp and Tenenbaum 2008; Rigotti et al. 2010; Gershman 2017; Russek et al. 2017). To improve experimental access to these phenomena, we sought behavioral evidence for interactions between learning at different hierarchical levels, namely, learning of individual objects and learning of the temporal context in which such objects are experienced.

Sequences of visual presentations may exhibit different kinds of temporal structure arising from sequential dependencies. A simple kind of structure is sequential dependency between consecutively presented items (i.e., an increased probability of item X, given preceding item Y). A more complex kind of structure arises when sequential dependencies are clustered within subsets of items. This leads to longer-term dependencies (i.e., an increased probability of item X, given recent item Z) and extended sequences of items that are mutually predictive (Schapiro et al. 2013; Karuza et al. 2017; Kahn et al. 2018).

The mechanisms of visual object recognition have been studied extensively (Wallis and Bülthoff 1999) with considerable evidence supporting "feature-based mechanisms" that represent three-dimensional objects in terms of multiple two-dimensional features/views (plus interpolations) (Bülthoff and Edelman 1992). Presumably, temporal regularities arise naturally in handling three-dimensional objects and help associate distinct two-dimensional views and/or features (Wallis and Bülthoff 1999). For example, when nonhuman primates learn to categorize initially unfamiliar objects, they readily form neural representations for arbitrary two-dimensional features that are diagnostic for category (Sigala and Logothetis 2002; Sigala et al. 2002). Interestingly, such representations automatically encompass predictive sequential dependencies between successive trials, even when its diagnostic information is redundant (Miyashita 1988; Wallis 1998).

The effect of sequential dependencies between successive trials on visual object recognition was investigated by two previous studies, which found a reaction time advantage (Barakat et al. 2013) and a recognition memory advantage (Otsuka and Saiki 2016) for target objects that consistently follow particular objects, compared with target objects that follow varying objects. Here we extended these findings in two ways: First, we monitored the formation of recognition memory more closely and comprehensively (every presentation of every object), and second, we considered the effect of clustered dependencies creating "temporal communities" of objects (which are typically experienced for nine successive presentations).

We investigated performance of observers in a visual object recognition learning task under three conditions: (1) "strongly structured" sequences comprising distinct temporal communities (clusters of mutually predictive objects), (2) "weakly structured" sequences with uniform sequential dependence, and (3) "random" or "unstructured" sequences without sequential dependence. All sequences were generated as random walks on graphs of $n = 15$ distinct objects (Fig. 1A), in which nodes represented distinct objects and edges represented possible transitions (in both directions). As one sequence comprised 180 object presentations, each graph was

Corresponding author: ehsankakaei91@gmail.com
Article is online at http://www.learnmem.org/cgi/doi/10.1101/lm.053306.120.

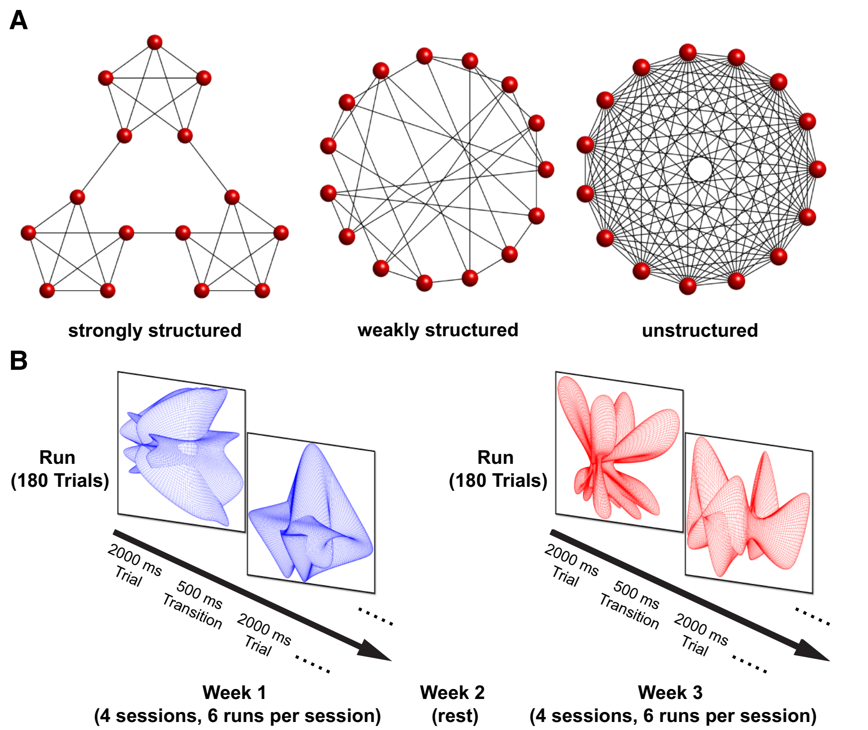

Figure 1. Presentation sequence and trial structure. (*A*) Presentation sequences were generated as (nearly) random walks on three types of graphs, with nodes representing a distinct object and edges representing possible transitions (in both directions). A sparsely connected, modular graph generated "strongly structured" sequences with distinct community structures (*left*), a sparsely connected, nonmodular graph generated "weakly structured" sequences (*middle*), and a full connected graph generated "unstructured" or "random" sequences (*right*). (*B*) Presentation sequences consisted of 180 complex, three-dimensional objects (shown rotating for 2 sec about a randomly oriented axis in the frontal plane). Of these, 170 ± 0.04 (mean ± SEM) objects were recurring, and 9.2 ± 0.04 objects were nonrecurring. Observers categorized each object as "familiar" or "unfamiliar." Over the four sessions of 1 wk, observers performed 24 runs and viewed 4320 presentations, with every recurring object appearing at least 250 times.

traversed multiple times (~11.3 times). Graphs were either modular and sparsely connected ("strongly structured" sequences), or nonmodular and sparsely connected ("weakly structured" sequences), or nonmodular and fully connected ("unstructured" or "random" sequences). In "strongly structured" sequences, approximately 9.2 ± 0.1 successive presentations (mean ± SEM) featured objects of the same temporal community.

One presentation sequence ("run") comprised exactly 180 objects and on average included 9.2 ± 0.04 (mean ± SEM) nonrecurring objects appearing exactly once during the entire experiment. Nonrecurring objects were spaced 14–19 presentations apart. The remaining 170 ± 0.04 objects were recurring and were selected by performing a pseudorandom walk on a graph (Fig. 1A), albeit with some restrictions: no direct repeats and returns were permitted (e.g., X–X or X–Y–X) and all $n = 15$ objects were repeated comparably often (11.4 ± 0.04 repetitions). The repetition latency for any given object ranged from three to >60 presentations. Very short latencies (of three to five presentations) were far more common in strongly structured sequences than in weakly structured or unstructured sequences (Supplemental Fig. S8).

To control the difficulty of shape recognition, ensure initial unfamiliarity of all objects, and minimize interference from semantic associations, we generated complex three-dimensional objects by convolving two closed Bezier curves in a plane. Complexity was controlled by number and the position of random seeds for the two curves. The pairwise dissimilarity of the resulting complex objects was statistically unrelated to their pairwise distance in the presentation sequence (see Supplemental Fig. S1). To ensure this, dissimilarity was quantified in terms of the vector distance between depth maps (of resolution $64 \times 64 \times 64$) obtained from six viewing directions along the three principal component axes.

Objects were presented for 2 sec rotating with an angular velocity of 144 deg/sec about an axis in the frontal plane. Starting angle and axis orientation were randomized for each trial, forcing observers to become familiar with the full three-dimensional shape (rather than just certain features). Presentation periods were separated by 0.5-sec transition periods, during which the previous object disappeared toward a distant location on the right, while the next object approached from a distant location on the left. This was intended to encourage observers to imagine a spatially extended sequence of distinct objects (Supplemental_Movie_S1).

Twenty healthy observers (eight males and 12 females, aged 25 to 34 yr old) participated in three experiments. Two experiments compared "strongly structured" and "unstructured" sequences, and one experiment compared "strongly structured" and "weakly structured" sequences. All observers had normal or corrected to normal vision and were paid for their participation. Ethical guidelines of the Centre for Neuroscientific Innovation and Technology, Magdeburg, were followed.

In order to monitor the progress of recognition learning as closely as possible, observers were required to classify every object presented as either "familiar" (seen previously) or "unfamiliar" (never seen previously). For each observer, a fresh set of 30 pairwise dissimilar objects was generated. The set was divided arbitrarily into two subsets of 15 objects, one used for "structured" sequences and the other for "unstructured" sequences. In addition, we generated a larger number (~500) of nonrecurring objects, which appeared exactly once during the entire experiment. During each trial, the observer categorized the current object as "familiar," "unfamiliar," and "not sure," by pressing a key. No feedback was provided. Observers performed this task on four different days within 1 wk, with six sequences per day (24 sequences overall). Accordingly, observers viewed 4320 presentations during which every recurring object appearing at least 250 times. After pausing for a week, observers repeated the experiment with entirely new objects and with sequences generated from another graph (Fig. 1B). Observers were told that each condition used new objects that were never shown before. To further emphasize this point, object color changed between conditions. The order of conditions (structured or unstructured) was counter-balanced between observers. Observer instructions did not mention presentation order (sequence structure).

At the end of each week of testing, observers were required to additionally perform a validation task, to assess the extent to which objects had become familiar (Supplemental_Movie_S2; Supplemental_Material). In this task, observers viewed for 30 sec an array of 12 simultaneously rotating objects, of which three were randomly selected from the 15 "recurring" objects and nine objects were entirely new (never seen before). Observers were asked to pick out the three most "familiar" objects and received binary feedback ("all correct" or "one or more incorrect"). All observers approached ceiling performance (proportion correct >0.95) in all conditions (all sequence structures), confirming that almost all recurring objects had become familiar.

To establish the progress of recognition learning, we analyzed 250 repetitions (over four sessions and 24 sequences) of every recurring object. To this end, we considered "sliding windows" with $N_w = 5$ successive presentations of a given object (for details see Supplemental Fig. S3). Note that some windows bridged successive presentation sequences and/or sessions. For each window and "recurring" object, we computed the proportion of "familiar" responses ("hit rate") (Fig. 2A). As "familiar" objects were common, some false positives were to be expected. To take this into account, we also established a "false alarm rate" for each session, as the fraction of "nonrecurring" objects not categorized as "unfamiliar" (Fig. 2B). Combining hit rate (of a window) with false alarm rate (of the concomitant session), we performed a simplified sensitivity analysis (Macmillan and Creelman 2004) to obtain a corrected classification performance ρ and decision bias b for each window and "familiar" object (see the Supplemental_Material). Alternative sensitivity analyses and performance measures (A', d'; Stanislaw and Todorov 1999) did not materially alter the results.

The resulting corrected performance ρ (mean and SEM, assuming binomial variability) is shown in Figure 2C. Performance increased nearly monotonically, but was consistently superior when objects were presented with "strongly structured" sequences with "temporal community structure" than when they were presented in unstructured sequences. This difference was significant after ~60 presentations. As expected, observers rapidly developed a liberal bias (favoring "familiar" responses), which weakened somewhat over subsequent sessions (Fig. 2D).

We also analyzed the time-development of average response times (RTs). Consistent with the performance results, RTs decreased faster for strongly structured sequences than for unstructured sequences (Supplemental_Material; Supplemental Fig. S2).

In addition to the gradual increase in the probability of recognizing recurring objects, we also sought to determine the point in time at which individual objects became familiar ("onset of familiarity"). We defined this point in two alternative ways: (1) as the first window in which corrected performance exceeded a threshold of $\rho \geq 0.875$ (high threshold approach) or (2) as the window in which entropy $H_\rho = -[\rho log_2(\rho) + (1-\rho) log_2(1-\rho)]$ of corrected performance reached its peak value (low threshold approach). Note that entropy peaks at the transition from exclusively "unfamiliar" to exclusively "familiar" responses.

After establishing the "onset of familiarity" for each object, we ranked all objects by order of onset and established the "onset separation" between object pairs in terms of onset rank (Δn) and presentation rank (Δk). The median separation of successive onsets (defined by threshold or entropy) was nine or 16 presentations, respectively. Interestingly, the median separation of successive onsets in same cluster was roughly thrice as long, with 24 and 50

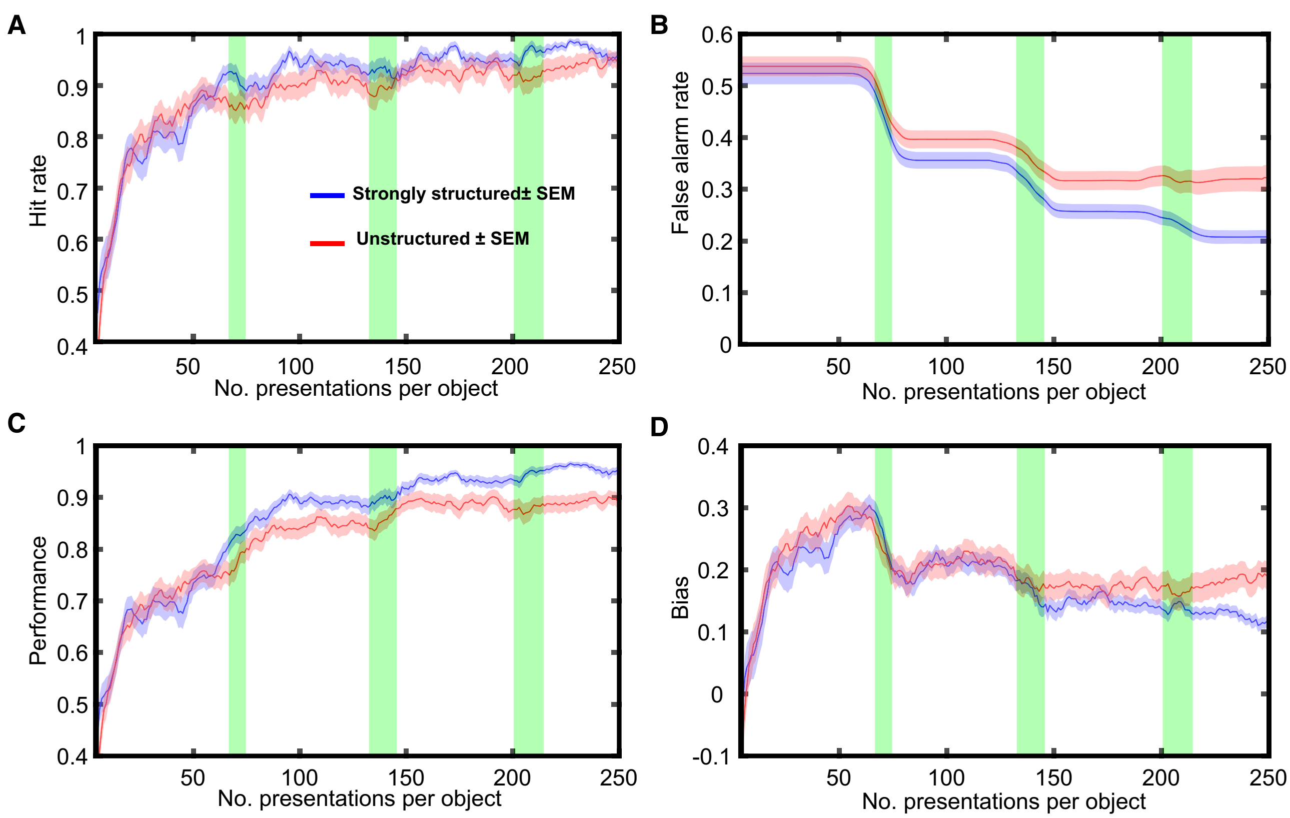

Figure 2. Time course of recognition learning. (*A*) Average hit rate (recurring categorized as familiar, per window) increases with the number of presentations of a given object. (*B*) Average false alarm rate (nonrecurring not categorized as unfamiliar, per session) decreases with the number of presentations. (*C*) Average corrected performance ρ increases nearly monotonically with presentation number. It was consistently larger for strongly structured sequences (with temporal community structure) than for unstructured sequences. (*D*) Average criterion bias *b*, as a function of presentation number. Green regions indicate the transition between sessions (20%–80% of objects in previous session).

presentations, respectively, implying that successive onsets occurred during separate visits to a given community.

In strongly structured sequences, one may distinguish objects pairs XY that are "adjacent" [follow each other with $P(Y|X) = 0.25$] or "nonadjacent" [never follow each other, $P(Y|X) = 0$]. In addition, one may distinguish object pairs within the same community (either adjacent or nonadjacent) and between different communities (also either adjacent or nonadjacent). Note that the objects linking different communities ("linking objects") contribute both "adjacent" pairs in different communities and "nonadjacent" pairs in the same community (Fig. 3B). We analyzed the "onset of familiarity" for different object pairs (as defined above), specifically, the probability that the two members of a pair exhibit successive onsets ($\Delta n = 1$) or nearly successive ($\Delta n = 2$) onsets. Interestingly, the probability of successive onsets was significantly higher than chance for objects in the same community (null hypothesis H_0: "onsets" are ordered randomly) (Fig. 3A). Moreover, we found the probability of successive "onsets" to be significantly elevated for "adjacent" objects in the same cluster, insignificantly elevated for "adjacent" objects in different clusters ("linking objects"), and significantly reduced for "nonadjacent" objects in different clusters ($P < 0.05$; corrected for false discovery rate of multiple comparisons) (Fig. 3B; Benjamini and Hochberg 1995).

We conclude that temporal community structure had a significant effect on the order of recognition learning in the sense that familiarity of one object in a community facilitated familiarity of another object in the same community, provided the latter was "adjacent" [i.e., followed the former sometimes, $P(Y|X) = 0.25$]. Interestingly, no such "domino effect" was observed for the objects linking two different communities (i.e. adjacent objects in different communities).

The results presented in Figures 2 and 3 were replicated with an additional eight observers in a second experiment of almost identical design (Supplemental Figs. S4, S6).

To dissociate the effects of cluster-membership and adjacency, we also conducted a third experiment, in which six further observers viewed either "weakly structured" presentation sequences (during 1 wk) or "strongly structured" sequences (during another week). To generate "weakly structured" sequences without temporal communities, we generated sparsely connected graphs with exactly four links per node, but without any triangular link formations (Maslov and Sneppen 2002; Rubinov and Sporns 2010). Recognition learning was faster for "strongly structured" sequences than for "weakly structured" ones. The "domino" effect described above was again observed for "strongly structured" sequences (with both "onset" definitions), but to some extent also for "weakly structured" sequences (for one "onset" definition). Thus, the ordering of "onsets" of familiarity may be affected both by community membership and by adjacency in the presentation sequence (Supplemental Figs. S5, S7).

In this study, we investigated the effect of temporal community structure by comparing more or less structured presentation sequences. First, in "weakly structured" sequences, sparse connectivity of the generative graph ensured that each object predicted the next object with 25% probability (one of four possibilities). Second, in "strongly structured" sequences, the (equally sparse) generative graph was clustered into three communities of five objects, so that each object predicted the community membership of the next object with 90% probability (18 of 20 possibilities).

Previous studies of statistical learning did not aim to closely follow the learning of individual items (Siegelman et al. 2018). Here we sought to monitor the degree of familiarity of each individual object over successive presentations (Fig. 2). Whereas classification performance improved monotonically with presentation number for all sequences, a significant performance advantage developed quickly (over 60 to 70 presentations) for "strongly structured" sequences compared with either "unstructured" or "weakly structured" sequences (Supplemental Fig. S5). Note that recognition performance improved comparably over time, with or without having practiced stimulus-response mapping in a separate training session (experiments 2 and 3). Accordingly, we do not believe that motor learning contributed appreciably to these results.

Thanks to close monitoring, we could almost always determine the onset of familiarity for an individual object. Interestingly, the ordering of onsets did not appear to be fully random, in that objects of the same community ("temporal community") tended to become familiar after one another more often than expected by chance. Interestingly, this "domino effect" typically did not occur within one "extended visit" to a community but over subsequent visits to a given community. This "domino effect" was particularly pronounced for adjacent objects in the same

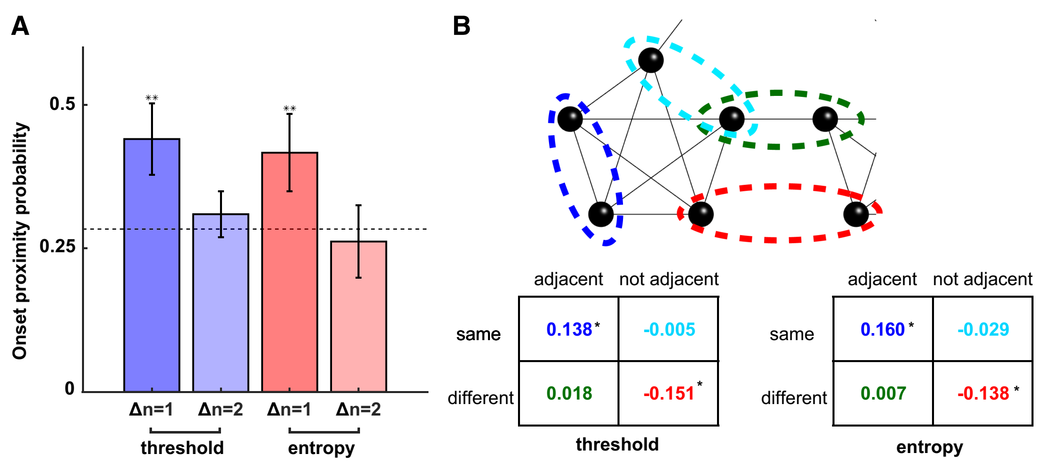

Figure 3. Analysis of the onset of familiarity with individual objects. (*A*) Successive onsets of familiarity ($\Delta n = 1$) are far more likely ([**] $P < 0.005$) for objects within the same cluster than would be expected by chance (dashed line). For nearly successive onsets ($\Delta n = 2$) this effect was not observed. (*B*) Comparison of frequency of successive onsets, compared with chance level, for objects pairs either in the same cluster (outlined blue and cyan) or in different clusters (green and red), which are either adjacent (blue and green) or nonadjacent on the graph (cyan and red). Frequency is significantly elevated ([*] $P < 0.05$ FDR corrected) for adjacent objects in the same cluster (blue) and suppressed for nonadjacent objects in different clusters (red).

community, but was not observed for adjacent objects in different communities. As a similar effect was observed for adjacent objects in "weakly structured" sequences without communities, there seems to be a contribution of frequent temporal proximity.

At the end of training, all objects had become familiar and could be retrieved explicitly from long-term memory, for both structured and unstructured sequences. The reason for the observed difference in learning rates remains unclear. One possibility is that structured sequences pose a reduced working-memory load, facilitating encoding and accelerating learning. When large sets of items are divided ("chunked") into subsets, both chunked and nonchunked items benefit and are learned more readily. Presumably, chunking reduces the dimensionality of the classification problem presented by each item (just like chunking the search array in an odd-man-out task reduces the dimensionality of target detection). This reduced dimensionality could then lower working-memory load and facilitate classification by comparison with long-term memory for both familiar (chunked) items and unfamiliar (nonchunked) items. Another important factor might be that temporal communities reduce repetition latencies (Supplemental Fig. S8). There is evidence that timely repetitions help consolidate memories, whereas delayed repetitions leave memories prone to disruption (Thalmann et al. 2019).

Previous studies of the effect of "temporal community structure" have shown that cluster borders are detectable (Schapiro et al. 2013) and that such borders elevate reaction time (Kahn et al. 2018; Karuza et al. 2019). As border items are thought to facilitate encoding/retrieval (Swallow et al. 2009), one might have expected accelerated recognition learning for "linking objects" that join two different clusters. However, in our paradigm, neither learning rate nor ordering of onsets of familiarity distinguished "linking objects" from other objects. In fact, our results suggest that any chunking benefits (Thalmann et al. 2019) apply more to objects within clusters than to objects that "link" clusters.

In summary, we showed that the presence of temporal communities of mutually predictive objects accelerates recognition learning for complex, three-dimensional objects and alters the order of recognition learning such that members of a group are often learned after one another (but separated by many intervening presentations).

Acknowledgments

The project was funded by the federal state Saxony-Anhalt and the European Structural and Investment Funds (ESF, 2014-2020), project number ZS/2016/08/80645.

References

Barakat BK, Seitz AR, Shams L. 2013. The effect of statistical learning on internal stimulus representations: predictable items are enhanced even when not predicted. *Cognition* **129:** 205–211. doi:10.1016/j.cognition.2013.07.003

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B (Methodol)* **57:** 289–300. doi:10.1111/j.2517-6161.1995.tb02031.x

Bülthoff HH, Edelman S. 1992. Psychophysical support for a two-dimensional view interpolation theory of object recognition. *Proc Natl Acad Sci* **89:** 60–64. doi:10.1073/pnas.89.1.60

Chun MM, Jiang Y. 1998. Contextual cueing: implicit learning and memory of visual context guides spatial attention. *Cogn Psychol* **36:** 28–71. doi:10.1006/cogp.1998.0681

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

Fiser J, Aslin RN. 2002. Statistical learning of higher-order temporal structure from visual shape sequences. *J Exp Psychol Learn Mem Cogn* **28:** 458–467. doi:10.1037/0278-7393.28.3.458

Gershman SJ. 2017. Predicting the past, remembering the future. *Curr Opin Behav Sci* **17:** 7–13. doi:10.1016/j.cobeha.2017.05.025

Hamid OH, Wendemuth A, Braun J. 2010. Temporal context and conditional associative learning. *BMC Neurosci* **11:** 45. doi:10.1186/1471-2202-11-45

Jiang Y, Wagner LC. 2004. What is learned in spatial contextual cuing: configuration or individual locations? *Percept Psychophys* **66:** 454–463. doi:10.3758/BF03194893

Kahn AE, Karuza EA, Vettel JM, Bassett DS. 2018. Network constraints on learnability of probabilistic motor sequences. *Nat Hum Behav* **2:** 936–947. doi:10.1038/s41562-018-0463-8

Karuza EA, Kahn AE, Thompson-Schill SL, Bassett DS. 2017. Process reveals structure: how a network is traversed mediates expectations about its architecture. *Sci Rep* **7:** 1–9. doi:10.1038/s41598-017-12876-5

Karuza EA, Kahn AE, Bassett DS. 2019. Human sensitivity to community structure is robust to topological variation. *Complexity* **2019:** 8379321. doi:10.1155/2019/8379321

Kemp C, Tenenbaum JB. 2008. The discovery of structural form. *Proc Natl Acad Sci* **105:** 10687–10692. doi:10.1073/pnas.0802631105

Macmillan NA, Creelman CD. 2004. *Detection theory: a user's guide*, 2nd ed. Lawrence Erlbaum Associates, Mahwah, NJ.

Maslov S, Sneppen K. 2002. Specificity and stability in topology of protein networks. *Science* **296:** 910–913. doi:10.1126/science.1065103

Miyashita Y. 1988. Neuronal correlate of visual associative long-term memory in the primate temporal cortex. *Nature* **335:** 817–820. doi:10.1038/335817a0

Otsuka S, Saiki J. 2016. Gift from statistical learning: visual statistical learning enhances memory for sequence elements and impairs memory for items that disrupt regularities. *Cognition* **147:** 113–126. doi:10.1016/j.cognition.2015.11.004

Rigotti M, ben Dayan Rubin D, Morrison SE, Salzman CD, Fusi S. 2010. Attractor concretion as a mechanism for the formation of context representations. *Neuroimage* **52:** 833–847. doi:10.1016/j.neuroimage.2010.01.047

Rubinov M, Sporns O. 2010. Complex network measures of brain connectivity: uses and interpretations. *Neuroimage* **52:** 1059–1069. doi:10.1016/j.neuroimage.2009.10.003

Russek EM, Momennejad I, Botvinick MM, Gershman SJ, Daw ND. 2017. Predictive representations can link model-based reinforcement learning to model-free mechanisms. *PLoS Comput Biol* **13:** e1005768. doi:10.1371/journal.pcbi.1005768

Saffran JR, Aslin RN, Newport EL. 1996. Statistical learning by 8-month-old infants. *Science* **274:** 1926–1928. doi:10.1126/science.274.5294.1926

Schapiro AC, Turk-Browne NB. 2015. Statistical learning. In *Brain mapping: an encyclopedic reference* (ed. Toga AW), pp. 501–506. Academic Press, New York.

Schapiro AC, Rogers TT, Cordova NI, Turk-Browne NB, Botvinick MM. 2013. Neural representations of events arise from temporal community structure. *Nat Neurosci* **16:** 486–492. doi:10.1038/nn.3331

Siegelman N, Bogaerts L, Kronenfeld O. 2018. Redefining 'learning' in statistical learning: what does an online measure reveal about the assimilation of visual regularities? *Cogn Sci* **42:** 692–727. doi:10.1111/cogs.12556

Sigala N, Logothetis NK. 2002. Visual categorization shapes feature selectivity in the primate temporal cortex. *Nature* **415:** 318–320. doi:10.1038/415318a

Sigala N, Gabbiani F, Logothetis NK. 2002. Visual categorization and object representation in monkeys and humans. *J Cogn Neurosci* **14:** 187–198. doi:10.1162/089892902317236830

Sisk CA, Remington RW, Jiang YV. 2019. Mechanisms of contextual cueing: a tutorial review. *Atten Percept Psychophys* **81:** 2571–2589. doi:10.3758/s13414-019-01832-2

Stanislaw H, Todorov N. 1999. Calculation of signal detection theory measures. *Behav Res Methods Instrum Comput* **31:** 137–149. doi:10.3758/BF03207704

Swallow KM, Zacks JM, Abrams RA. 2009. Event boundaries in perception affect memory encoding and updating. *J Exp Psychol Gen* **138:** 236–257. doi:10.1037/a0015631

Thalmann M, Souza AS, Oberauer K. 2019. How does chunking help working memory? *J Exp Psychol Learn Mem Cogn* **45:** 37–55. doi:10.1037/xlm0000578

Wallis G. 1998. Temporal order in human object recognition learning. *J Biol Syst* **6:** 299–313. doi:10.1142/S0218339098000200

Wallis G, Bülthoff H. 1999. Learning to recognize objects. *Trends Cogn Sci* **3:** 22–31. doi:10.1016/s1364-6613(98)01261-3

Received November 24, 2020; accepted in revised form February 13, 2021.