What is Represented in Memory after Statistical Learning?

Tess Allegra Forest (tess.forest@mail.utoronto.ca), Amy S. Finn, and Margaret L. Schlichting
Department of Psychology, University of Toronto
100 St. George Street, ON, M5S 3G3

Abstract

Statistical learning is a powerful mechanism that allows us to rapidly extract structure from the environment. However, nuances of what structure is extracted—for example, whether reliable groups are stored without knowledge of their constituent item order—are not well understood, leaving us with open questions about how this mechanism supports behaviour. Here, we extend prior work on the representation of statistical structure by asking what specific aspects of structure matter for memory judgments. We consider three candidates for memory representation: transitional probability, order-independent group information, and position tags. Participants watched a stream of shape triplets and then completed a recognition memory test designed to isolate contributions of transitional probability, group, and position. We demonstrate that although memory for transitions alone would be sufficient for knowledge of triplets, participants showed evidence of representing both transitional probability and group. Our data highlight statistical learning as a mechanism enabling generalization across experiences.

Keywords: Statistical Learning; Memory

Introduction

Our world is full of structure—structure we are well-equipped to learn about via statistical learning (Saffran, Aslin, & Newport, 1996). Extracting regularities from experience is both broadly *beneficial* to behaviour—helping us segment words from continuous speech and learn about object co-occurrences—and broadly *available* across ages and species (Santolin and Saffran, 2018).

Prior work suggests that statistical learning operates by computing the predictive relationships between individual items (for example, "A always precedes B"), ultimately yielding knowledge of multi-item groupings through concatenation of these specific item-item relationships (Saffran et al., 1996). This dominant perspective makes clear predictions about what is stored in memory—and thus, what should feel familiar upon later encounters. In particular, familiarity should reflect the consistency between item-item transitions and previous experience. Intuitively, however, there are also other statistics that a learner could store in memory.

Here, we simultaneously assessed three possibilities for what shapes memory when there are multiple statistics a learner could use. First, we considered memory for transitional probability, reflecting specific item-item transitions. Second, we assessed memory for group membership, in which commonalities across presentations of one triplet are emphasized at the expense of specific transitions, yielding, at its most extreme, order-irrelevant memory for groups (Gilboa & Marlatte, 2017; Schlichting,

Mumford, & Preston, 2015). Finally, we measured memory for ordinal position, in which items are "tagged" with their ordinal within-group position (Hsieh, Gruber, Jenkins, & Ranganath, 2014; Kikumoto & Mayr, 2018).

Order independent group membership information aligns with another historical perspective on how statistical learning proceeds, typically referred to as "chunking". This school of thought has suggested that instead of tracking the specific, predictive relationships between items, associations between individual items are learned *without* learning the predictive relationships between them (Endress & Mehler, 2009; Perruchet & Vinter, 1998). Thus, this perspective suggests the representations formed as a result of statistical learning only include group level information about which items go together reliably. However, this perspective has not made strong claims about whether the order of items in those groups is maintained (Perruchet & Vinter, 1998). Consequently, most of this previous work comparing transitional probability-based to "chunk"-based representations have not manipulated the order of items relative to exposure (e.g. Endress & Mehler, 2009), making it unknown whether item order is in fact stored.

Here, we were specifically interested in whether *order-independent* group information is stored in memory following statistical learning—a question left unanswered by the existing chunking literature. That said, other past work has suggested that order-independent representations can be acquired via statistical learning; however, this has thus far only been observed when other signals (e.g., reliable transitions) are intentionally absent from the input (Karuza, Kahn, Thompson-Schill, & Bassett, 2017; Schapiro, Rogers, Cordova, Turk-Browne, & Botvinick, 2013). Thus, it remains unclear whether order-independent group representations will emerge despite exposure to a standard statistical stream—that is, when learning transitions would be sufficient to support behaviour.

Our final possibility for what shapes memory was ordinal position. Past work from sequence learning tasks suggests that ordinal position is coded in memory to guide eye-movements (Pathman & Ghetti, 2015), and that this information is neurally represented alongside the identity of the broader sequence (Kikumoto & Mayr, 2018). There is also neural evidence for position-specific coding in the parahippocampal cortex (Hsieh et al., 2014). This provides a mechanism by which position may also be coded separately from item-item transitions in statistical learning tasks. However, whether position similarly influences memory in a behavioural statistical learning paradigm remains unknown.

To ask whether any or all of these possibilities are represented in memory after statistical learning, we

systematically manipulated the degree to which probe sequences during a memory test matched exposure along these dimensions and quantified their respective impacts on behaviour.

Methods

Participants

Data were collected online (https://prolific.ac/) after preregistration (https://osf.io/gjt8x). One hundred native-English speakers between the ages of 19 and 35 who reported no history of neurological, psychiatric, or learning disorders were included in the final sample (Mean age = 27.84 years, SD = 4.68, 28 female). An additional 33 participants were excluded due to lack of attention during exposure, which was defined as failure to respond to one or more attention checks (described below). All procedures were approved by the Research Ethics Board at the University of Toronto. Previous studies of statistical learning and memory (Batterink, Reber, Neville, & Paller, 2015) have reported effect sizes in the medium to large range (d = 0.60 used for power calculation). Power calculations suggested a sample size of 66 subjects was required for 80% power in a multiple regression with three predictors. However, because of our manipulation of correlated factors and results of a pilot study, we anticipated our effects to be slightly smaller. Thus, we increased the sample size to 100 to ensure enough participants to avoid Type I error.

Stimuli

All stimuli were presented to participants on their own computers using Inquisit 5.0.11 (2018, https://millisecond.com). Stimuli were nine shapes, each of a unique bright color. They were presented one at a time in the centre of the screen on a black background in a continuous 'stream.' Unbeknownst to participants, there was a particular structure built into the shape stream: Shapes were organized into three 'triplets,' or groups of three shapes that always appeared in the same, fixed order. For example, given the three triplets ABC, DEF, and GHI, shape A would always precede shape B, which would in turn always precede shape C. Triplet ABC could then be followed either by the triplet DEF or GHI. This relatively small number of items was chosen to ensure that this paradigm could be used with children in future experiments. Triplets were matched in average luminance and RGB values of the shapes and were held constant across participants, similar to other previous studies of visual statistical learning (see Kirkham, Slemmer, & Johnson, 2002).

Task Design

Exposure Participants watched the stream of triplets for a total of 6.5 minutes. Shapes were presented for 1000 ms, with 100 ms inter-stimulus interval [ISI]. Each triplet was repeated 30 times, and triplets could appear in any order with the caveats that first, no triplet was allowed to immediately follow itself and second, a given triplet was followed by each of the other two triplets exactly 50% of the time (e.g. Saffran, Aslin, & Newport, 1996). Transitional probability (TP), or the probability that a particular item (Y) will appear next given the current item (X), therefore varied based on whether a transition from X to Y was a within-triplet transition (e.g., TP_{within} = 1.0) or was a between-triplet transition ($TP_{between}$ = .5). Triplets were consistent between all participants.

As this experiment was run remotely, we included three presentations of an attention check trial that allowed us to identify participants who were not complying with task instructions to '*please simply watch the images.*' In these attention checks, participants saw the text '*Please press "F" as quickly as possible*' and had 1.5 seconds to make their keypress response. Participants who failed one or more of the attention checks were excluded from the final sample and a new participant was run in their stead. These trials were included purely for purposes of participant exclusion and were otherwise unrelated to the task. Importantly, attention checks could not have been used as a cue for successful segmentation because they could appear both within and between triplets, with specific positions determined randomly for each participant.

Test Phase After exposure, participants completed a recognition memory test in which they indicated whether they thought a *sequence* of three shapes was part of what they had watched during exposure (old) or if it was not (new). Three of the sequences were truly old (sequences they had seen previously) and 17 were new (detailed below). Because all shapes were seen at exposure, participants' judgements about test sequences could not be based on familiarity with the shapes themselves, but rather had to be made on the particular sequence of shapes. All test sequences were presented with each item appearing alone for 1000 ms with a 100 ms ISI, thus matching presentation timing during exposure. Each test sequence was repeated three times (for a total of 60 test trials) in a pseudorandom order such that repetitions were evenly distributed across thirds of the test.

Test Sequences The goal of this experiment was to understand what aspects of the exposure stream are represented in memory after learning. As such, we generated 20 test sequences that varied in the degree to which they corresponded with the exposure stream along three dimensions: 1) the sum total transitional probability (TP) across both forward transitions (the transition from the 1st item to the 2nd, and the 2nd to the 3rd) in a test sequence (hereafter 'TP'); 2) the number of shapes in a test sequence that had been part of the same exposure triplet ('Triplet Membership Score'), and 3) the total number of shapes which maintained their position within a test sequence relative to their exposure triplet (1st, 2nd, 3rd ; 'Position Score'). The specific sequences were generated so as to orthogonalize the

Table 1: Sample Test Sequences (for one participant)

Test Type	Test Sequence	Position Score	Transitional Probability	Triplet Membership
Exposure Triplets	ABC	3	2	3
	DEF	3	2	3
	GHI	3	2	3
Novel Test Sequences	ACB	1	0	3
	BAC	1	0	3
	EFD	0	1	3
	FED	1	0	3
	IGH	0	1	3
	GBF	3	0	1
	AIE	0	0	1
	EIA	0	0.5	1
	HDC	1	0	1
	CDH	0	0.5	1
	FBG	1	0	1
	AHC	3	0	2
	DIH	1	0	2
	BGI	1	0	2
	EFG	0	1.5	2
	CED	1	0	2
	FAB	0	1.5	2

relative contributions of these three dimensions as much as possible (*Table 1*). This test sequence generation process was carried out as follows: individual shapes were distributed evenly across test sequences such that no participant was exposed to any individual items more frequently than others during test. Moreover, for all test sequences that were not exposure triplets, specific items were evenly distributed across test sequences. Three counterbalanced orders were then generated to ensure that across participants, each of the test sequences were created from each triplet equally often.

To create the specific test sequences for one counterbalancing group, we first generated a list of all possible orderings of positions 1, 2, and 3 (of which there are six) at each level of triplet membership (i.e., 1, 2, or 3 items from the tested triplet; there are three levels). This yielded a total of 18 combinations (6 orderings at each of 3 triplet membership levels), one of which is the correct combination for an exposure triplet (Positions 1-2-3, with 3 items from the same triplet; *Table 1*). Because we tested each participant on all three exposure triplets (instead of just one), this resulted in a total of 20 test sequences per participant. Of these 20, three were original exposure triplets (meaning they were a sequence of three shapes which had always appeared together in the same order during exposure: ABC (Position 1-2-3), DEF (1-2-3), and GHI (1-2-3)); five were the remaining sequences which included three shapes from the same exposure triplet, but in a shuffled order (for example: BAC

(2-1-3) or IGH (3-1-2)); six were sequences with two shapes from one exposure triplet (for example: AEC (1-2-3) or IDE (3-1-2)); and six were sequences with one shape from each exposure triplet, akin to the 'position matched foils' commonly used in visual statistical learning paradigms (for example: AEI (1-2-3) or position matched to a shuffled item: IGB (3-1-2)) (see Park, Rogers, & Vickery, 2018).

Finally, we calculated the total TP for each sequence. For all sequences with a triplet membership score of 1 or 3, TP was already dictated by the position ordering. In test sequences with a triplet membership score of 2, any pair of shapes could have come from the same exposure triplet. Because of this greater flexibility in the choice of shapes (and thus TP) for the test sequences with a triplet membership score of 2, we split these sequences into two groups with different TP values to align our test sequences with prior literature. Group one included position orderings 3-1-2 and 2-3-1. For these sequences, the two same-triplet shapes selected created a sequence equivalent to 'part-word foils' used previously in the literature (FAB or BCD (Saffran et al., 1996)), and thus had a total TP of 1.5. The other four position orderings formed group two—here, the two same-triplet shapes selected created a test sequence with a total TP of 0 in order to most closely match their position ordering counterparts with triplet membership scores of 1 and 3 (for example AEC). Collectively, we chose our test sequences to

maximize our ability to isolate the separate effects of TP, triplet membership score, and position score.

Data Analysis

Considering each factor in isolation, TP and triplet membership score both showed a positive relationship with the tendency to endorse a test sequence as old (simple linear models; TP: $\beta = 0.15$, $SE = 0.009$, $z = 17.19$, $p < .0005$; triplet membership: $\beta = 0.10$, $SE = 0.008$, $z = 13.06$, $p < .0005$), whereas position showed a negative relationship (position score: $\beta = 0.03$, $SE = 0.006$, $z = 5.13$, $p < .0005$). However, these simple linear models will not capture the independent contributions of each factor, as there is inherent interdependence between factors (e.g., when position score is high, TP must be low). Thus, we chose our analysis approach to isolate the respective contributions of TP, triplet membership score, and position score to old-new judgments at test, after statistically controlling for the other factors.

Specifically, we ran a series of pre-registered generalized mixed effects models where we investigated the degree to which the composite scores—that is, how much each test sequence overall corresponded with what was seen during exposure on our three dimensions of interest—were associated with the probability of making an old response (see *Composite Score Effects*). We were additionally interested in whether any of the observed relationships were driven by specific transitions or positions being maintained between exposure and a test sequence. These questions were tested in two additional pre-registered models which broke down these two composite scores into their component parts (see *Specific Sub-Component Effects*).

Confirming Learning of Exposure Triplets First, to confirm that participants as a group had learned the triplets present in the exposure stream, the proportion of old responses to exposure triplets was compared to all other test sequences combined using a paired t-test. Successful learning across the group was defined as a greater proportion of old responses to old than new items.

Composite Score Effects To address our main question of which factors contribute to memory after controlling for the other possibilities, we constructed two generalized mixed-effects models (logistic regression) using the glmer function in the lme4 package (Bates, Machler, Bolker, & Walker, 2015; R Core Team, 2015) in R version 3.5.1. We ran two models at different levels of complexity to ask which better accounted for the data, which we refer to collectively as 'composite score effects models': first, a main effects model and second, a model that included a subset of all possible two-way interaction terms. These are referred to as composite score models because for each factor of interest (TP, triplet membership score, and position score), we calculated the total, composite score across the whole test sequence to result in one value on each dimension per test sequence (in contrast to the sub-component models described below). These scores were then treated as continuous numeric predictors. Both models included the main effects of TP, triplet membership score, and position score as fixed effects. The second model additionally included interactions between triplet membership score and position score and between triplet membership score and TP (see below). Both models also included random by-participant slopes and intercepts for main effects and interactions.

The interaction between triplet membership score and TP was included in the second model in order to account for the possibility that the degree to which TP informed old-new decisions would depend on the number of items from an exposure triplet present in the test sequence (triplet membership score). For example, raw TP between the individual shapes might matter more for test sequences which had a triplet membership score of 1—for which TP is the only cue to the oldness of the sequence—than it would in the context of a test sequence which had a triplet membership score of 3. This might occur because in a sequence with a triplet membership score of 1, intact transitions might be especially likely to underlie any erroneous 'old' responses. Similarly, the interaction between triplet membership score and position was included in the second model to account for the possibility that position score might be more important in test sequences with higher triplet membership scores. For example, the position of each shape might be more meaningful in a test sequence with a triplet membership score of 2 than in one with a score of 1, in which none of the shapes had occurred in the same triplet during exposure.

We compared the models with only main effects with the models including interactions using a BIC-based model comparison, which penalizes an overly complex model. Model comparison suggested the interaction model better accounted for the data (BIC Main Effects Model = 6644.3, BIC Interaction Model = 6627.7, $X^2(4) = 24.55$, $p < .001$). Thus, all the results presented below are from this better-fitting model which included the interaction terms. However, the directionality and overall significance of all three main effects was similar between the two models.

While TP, position score, and triplet membership score had a relatively low correlation (all r's < 0.4), test sequences that were exposure triplets (i.e., truly 'old' sequences) represented the highest value across all three dimensions and thus populated an extreme part of our test sequence space. Because we are interested in how subtle changes along each of our dimensions influence old-new judgements, rather than broad, categorical differences in how participants respond to old vs new sequences, we ran all our models on test sequences excluding truly old triplets. Thus, any results can be attributed fully to the effect that each dimension of interest has on falsely endorsing a new sequence as old. In other words, by removing these truly 'old' sequences, we can confirm that the exposure triplets were not driving any significant effects.

Specific sub-component effects To address the possibility that specific transitions or positions *within* a test sequence were particularly important for memory, we ran two follow-

up models that broke down our composite scores for TP and position into more specific elements. For each of these models, the other two main factors (e.g., triplet membership score and TP for the position model) were not included at all. Sub-component models were run regardless of the significance of the effects from the main models, as per our preregistered data analysis plan.

The first of these models was run to assess whether, ignoring the effects of position score or triplet membership score, a high probability first transition was more impactful for old-new judgements than a high-probability second transition. This model also included random, by-participant slopes and intercepts for these variables.

The second sub-component model was designed to assess whether, ignoring any effects of triplet membership score or TP, retaining an item in any *particular* position from exposure would be especially relevant for old-new judgements. For example, maintaining a Position 1 item from exposure in a position one spot in a test sequence could impact old-new judgements differently than maintaining a Position 2 or 3 item. This model included separate fixed effects for whether, irrespective of triplet, a test sequence's Position 1, 2, and 3 shape were in the same position (i.e., Position 1, 2, and 3 respectively) as they were during exposure as well as random, by-participant slopes and intercepts for these three variables.

There was no equivalent third follow-up model for the sub-components of our triplet membership score, because this composite score was simply an addition of the number of items that had been part of the same exposure triplet and thus did not have sub-components that could be sensibly separated.

Both transition and position sub-component models were run on all test items excluding truly 'old' exposure triplets, in order to confirm that the exposure triplets were not driving any significant effects, as with the main models.

Reporting Effect Sizes For all models in this study, results will be reported with Odds Ratios (1:x) as a measure of effect size appropriate for logistic regressions. Here, odds ratios imply that for every one-unit increase in the independent variable of interest (the left side of the ratio), there is an x-fold increase in the probability of a test sequence being endorsed as old.

Results

Results of a paired-samples t-test suggested that participants endorsed exposure triplets as old significantly more than all other test sequences (mean proportion old for exposure triplets = 0.68, mean proportion old for all other test sequences = 0.41, $t(99) = 8.33$, $p < .001$, $d = 1.25$), indicating that our measure of statistical learning was effective at capturing overall knowledge of which shapes appeared together reliably during exposure.

Composite Score Results: Transitional probability and triplet membership score impact old-new judgements Our main model asked how a test sequence's overall TP, triplet membership score, and position score impacted whether it was endorsed as old or new. The results of this analysis suggested that both TP and triplet membership score had large and reliable effects on participants' old-new judgements (TP: $\beta = 0.70$, $SE = 0.20$, $z = 3.56$, $p = .0004$, Odds Ratio = 1: 2.01, *Figure 2*; triplet membership score: $\beta = 0.33$, $SE = 0.13$, $z = 2.51$, $p = 0.01$, Odds Ratio = 1:1.39, *Figure 2*). In other words, participants were more likely to make old responses as a function of the number of both 1) transitions held intact and 2) shapes that occurred in a triplet together at exposure. Position score had no effect on the likelihood of a test sequence being endorsed as old ($\beta = -0.11$, $SE = 0.09$, $z = -1.24$, $p = 0.21$, Odds Ratio = 1:0.90, *Figure 2*), meaning that maintaining specific items in their position from exposure had no measurable effect on memory judgements.

Our model also included interaction terms, which allowed us to ask whether there were dependencies between our factors. In this model run on all data (excluding exposure triplets), the interaction between triplet membership score and TP was not significant ($\beta = 0.002$, $SE = 0.11$, $z = 0.03$, $p = 0.98$, Odds Ratio = 1:1.002). The interaction between triplet membership score and position score was also not significant ($\beta = 0.00$, $SE = 0.11$, $z = 0.03$, $p = 0.98$, Odds Ratio = 1:1.00).

Specific Sub-Component Results: Early positions may impact old-new judgements differently than later positions We next wanted to know whether specific *within-triplet* transitions or positions might be particularly important for memory judgements. Results from our two sub-component models indicated that, TP in the first and second transitions did not differentially impact old-new judgements (first transition: $\beta = 0.06$, $SE = 0.11$, $z = 0.54$, $p = 0.59$, Odds Ratio = 1:1.06; second transition: $\beta = -0.05$, $SE = 0.11$, $z = -0.49$, $p = 0.63$, Odds Ratio = 1:0.95), suggesting that each transition was equally likely to contribute to a participant erroneously endorsing a test sequence as 'old'.

As described above, composite position score—or how many items were maintained in a constant position exposure and test—did not predict old-new judgements in the main composite score model that controlled for the effects of TP and triplet membership score. However, when we assessed the effect of each position (i.e., 'slot' within a test sequence) individually ignoring the other two factors, maintaining the positions of *any* item from a triplet at exposure significantly decreased the likelihood of an 'old' judgment (first position: $\beta = -0.42$, $SE = 0.09$, $z = -4.76$, $p < .001$, Odds Ratio = 1:0.66; second position: $\beta = -0.28$, $SE = 0.10$, $z = -2.81$, $p = .005$, Odds Ratio = 1:0.76; third position: $\beta = -0.24$, $SE = 0.11$, $z = -2.24$, $p = 0.03$, Odds Ratio = 1:0.79), implying that position information—in contrast to the other two factors—may be informing participants' old-new decisions by serving to help them successfully reject new test sequences (i.e., correctly identifying them as 'new').
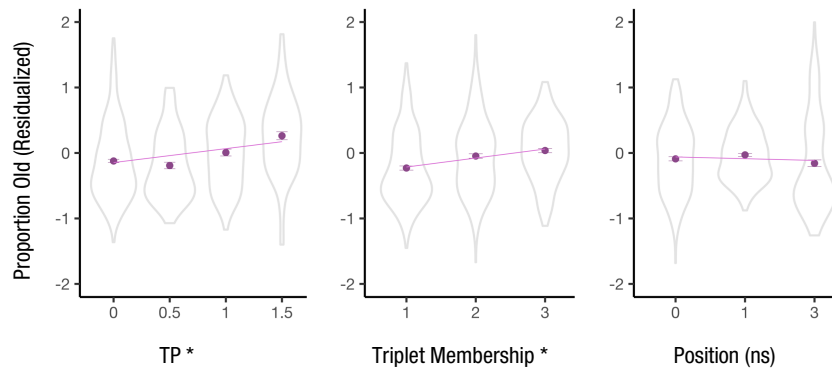
Figure 1: Residualized proportion of old responses new test sequences as a function of their total transitional probability, triplet membership score, and position score. Distribution of participant responses at each level of the predictor dimension (grey violins), along with mean values across participants for exposure sequences are shown (purple). Lines are for visualization purposes only and are derived from mixed models which predict the residualized old-new scores for each factor after controlling for every other factor and those interactions that included the predictor of interest. Error bars around the means and lines represent 95% confidence intervals.

Discussion

We show that memory judgements after statistical learning reflect both item-item TP *and* order-independent groupings, but not position. To our knowledge, this is the first demonstration that order-independent group information uniquely influences behaviour despite reliable item-item transition information being present in the input.

Test sequences more consistent with exposure TP were more likely to be endorsed as old, suggesting item-item links that unfold over time are stored in memory. This finding converges with the notion of *prediction* as one of memory's core functions (Norman & O'Reilly, 2003); indeed, brain regions implicated in memory—thought to link elements of experience as well as predict upcoming events—are critical for normal statistical learning (Schapiro, Turk-Browne, Norman, & Botvinick, 2016). We show old-new memory decisions reflect these links, building on the past literature which has largely measured *relative* oldness using forced choice tests.

Old-new decisions also tracked with the number of same-triplet items when controlling for other factors, highlighting another core memory function at play during statistical learning: combining across events to extract higher-order regularities. While past literature studying chunk-based representations largely left the order of items intact, here we shuffled the items within a triplet. As such, we conclude that representations formed as a result of statistical learning go beyond a set of item-item links to include integrated, order-independent group information. While perhaps surprising in the context of statistical learning studies emphasizing TP, this result converges with much episodic memory research. In particular, memory processes are important for segmenting experience into discrete events (DuBrow & Davachi, 2013) and linking reliably co-occurring elements (Schapiro, Kustner, & Turk-Browne, 2012)—even if such co-occurrences are indirect (Luo & Zhao, 2018; Schlichting et al., 2015). This finding also builds upon existing work demonstrating neural evidence for group representations (Schapiro et al., 2013) by demonstrating that this happens even when the structure is perfectly predictive. We propose a similar mechanism is at play in statistical learning, yielding a general triplet representation that impacts behaviour.

We saw no evidence that, when controlling for other factors, position influences behaviour. Interestingly, when position was examined on its own, maintaining each position had a negative impact on endorsing a new sequence as old. While the overall finding that position is not as important for old-new decisions as TP or triplet membership might seem at odds with prior work showing position coding in sequence memory tasks (Hsieh et al., 2014; Pathman & Ghetti, 2015), one important nuance of our task is that position is initially ambiguous and becomes evident only after stream segmentation. In other words, triplets must first be detected—like extracting events from a narrative (Kurby & Zacks, 2008). Consistent with research on event extraction, we found new test sequences with a 1st position item in position 1 were more often correctly judged as new, suggesting this position is particularly diagnostic for decisions as it 1) produces expectations for upcoming items and 2) serves as a boundary—especially important for memory (DuBrow & Davachi, 2013; Howard & Kahana, 2002).

Here, we provide both novel insight into past statistical learning studies and an explicit link to the memory literature. Our results also inform the design of future studies by providing a way to determine what factors influence behaviour after statistical learning. With this knowledge, we are better armed to address questions that consume cognitive neuroscientists and developmental psychologists alike: what sorts of information do learners of all ages extract and store from experience?

References

Bates, D., Machler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Batterink, L. J., Reber, P. J., Neville, H. J., & Paller, K. A. (2015). Implicit and explicit contributions to statistical learning. *Journal of Memory and Language*, *83*, 62–78. https://doi.org/10.1016/j.jml.2015.04.004

DuBrow, S., & Davachi, L. (2013). The influence of context boundaries on memory for the sequential order of events. *Journal of Experimental Psychology: General*, *142*(4), 1277–1286. https://doi.org/10.1037/a0034024

Endress, A. D., & Mehler, J. (2009). The surprising power of statistical learning: When fragment knowledge leads to false memories of unheard words. *Journal of Memory and Language*, *60*(3), 351–367. https://doi.org/10.1016/j.jml.2008.10.003

Gilboa, A., & Marlatte, H. (2017). Neurobiology of Schemas and Schema-Mediated Memory. *Trends in Cognitive Sciences*, *21*(8), 618–631. https://doi.org/10.1016/j.tics.2017.04.013

Howard, M. W., & Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology*, *46*(3), 269–299. https://doi.org/10.1006/jmps.2001.1388

Hsieh, L.-T., Gruber, M. J., Jenkins, L. J., & Ranganath, C. (2014). Hippocampal Activity Patterns Carry Information about Objects in Temporal Context. *Neuron*, *81*(5), 1165–1178. https://doi.org/10.1016/j.neuron.2014.01.015.Hippocampal

Karuza, E. A., Kahn, A. E., Thompson-Schill, S. L., & Bassett, D. S. (2017). Process reveals structure: How a network is traversed mediates expectations about its architecture. *Scientific Reports*, *7*(1), 1–9. https://doi.org/10.1038/s41598-017-12876-5

Kikumoto, A., & Mayr, U. (2018). Decoding hierarchical control of sequential behavior in oscillatory EEG activity. *ELife*, *7*, 1–36. https://doi.org/10.7554/eLife.38550

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition*, *83*(2), 4–5. https://doi.org/10.1016/S0010-0277(02)00004-5

Kurby, C. A., & Zacks, J. M. (2008). Segmentation in the perception and memory of events. *Trends in Cognitive Sciences*, *12*(2), 72–79. https://doi.org/10.1016/j.tics.2007.11.004

Luo, Y., & Zhao, J. (2018). Statistical Learning Creates Novel Object Associations via Transitive Relations. *Psychological Science*, *29*(8), 1207–1220. https://doi.org/10.1177/0956797618762400

Norman, K. A., & O'Reilly, R. C. (2003). Modeling Hippocampal and Neocortical Contributions to Recognition Memory: A Complementary-Learning-Systems Approach. *Psychological Review*, *110*(4), 611–646. https://doi.org/10.1037/0033-295X.110.4.611

Park, S. H., Rogers, L. L., & Vickery, T. J. (2018). The roles of order, distance, and interstitial items in temporal visual statistical learning. *Attention, Perception, and Psychophysics*, *80*(6), 1409–1419. https://doi.org/10.3758/s13414-018-1556-1

Pathman, T., & Ghetti, S. (2015). Eye movements provide an index of veridical memory for temporal order. *PLoS ONE*, *10*(5), 1–17. https://doi.org/10.1371/journal.pone.0125648

Perruchet, P., & Vinter, A. (1998). PARSER: A Model for Word Segmentation. *Journal of Memory and Language*, *39*, 246–263. Retrieved from papers://b4c16e83-3d5e-4b42-a8fa-32223867dbdf/Paper/p357

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928. https://doi.org/10.1126/science.274.5294.1926

Santolin, C., & Saffran, J. R. (2018). Constraints on Statistical Learning Across Species. *Trends in Cognitive Sciences*, *22*(1), 52–63. https://doi.org/10.1016/j.tics.2017.10.003

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*, *22*(17), 1622–1627. https://doi.org/10.1016/j.cub.2012.06.056

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. https://doi.org/10.1038/nn.3331

Schapiro, A. C., Turk-Browne, N. B., Norman, K. A., & Botvinick, M. M. (2016). Statistical learning of temporal community structure in the hippocampus. *Hippocampus*, *26*(1), 3–8. https://doi.org/10.1002/hipo.22523

Schlichting, M. L., Mumford, J. A., & Preston, A. R. (2015). Learning-related representational changes reveal dissociable integration and separation signatures in the hippocampus and prefrontal cortex. *Nature Communications*, *6*, 1–10. https://doi.org/10.1038/ncomms9151