AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology: General

ISSN: 0096-3445

2022, Vol. 151, No. 4, 837–851
https://doi.org/10.1037/xge0001104

General Precedes Specific in Memory Representations for Structured Experience

Tess Allegra Forest, Amy S. Finn, and Margaret L. Schlichting
Department of Psychology, University of Toronto, Saint George Campus

Decades of work has shown that learners rapidly extract structure from their environment, later leveraging their knowledge of what is more versus less consistent with prior experience to guide behavior. However, open questions remain about exactly what is remembered after exposure to structure. Memory for specific associations—transitions that unfold over time—is considered a prime candidate for guiding behavior. However, other factors could influence behavior, such as memory for general features like reliable groupings or within-group positions. We also do not yet know whether memory depends upon the amount of experience with the input structure, leaving us with an incomplete understanding of how statistical learning supports behavior. In 4 experiments, we tracked the emergence of memory for item-item transitions, order-independent groups, and positions by having 400 adults watch a stream of shape triplets followed by a recognition memory test. We manipulated how closely test sequences corresponded to the input along each dimension of interest, allowing us to isolate the contribution of each factor. Both item-item transitions and order-independent group information influenced behavior, highlighting statistical learning as a mechanism through which we form both specific and generalized representations. Moreover, these factors drove behavior after different amounts of experience: With limited exposure, only group information impacted old-new judgments specific transitions gained importance later. Our findings suggest statistical learning proceeds by first forming a general representation of structure, with memory being later refined to include specifics after more experience.

Keywords: statistical learning, memory representations, learning

Supplemental materials: https://doi.org/10.1037/xge0001104.supp

Our world is full of structure—structure we are well-suited to learn about via "statistical learning" (Saffran et al., 1996). Extracting regularities from experience is thought to support a broad range of behaviors, from helping us segment words in continuous speech to learning which objects reliably co-occur. While past research has suggested that we might extract regularities by either learning pairwise predictive relationships (Aslin et al., 1998) or item groupings (Perruchet & Vinter, 1998; see also Siegelman et al., 2019), these two options have largely been pitted against one another. Thus, as a field, we lack a clear understanding of how multiple representations may be formed through statistical learning to simultaneously influence behavior. What's more, we have very little information about how differences in the amount of experience with structured input a learner has change what they remember. The present study seeks to address these gaps in knowledge.

Our limited understanding about whether learners store multiple memory codes to guide their behavior after statistical learning is due in part to the way the mechanism has typically been studied. In a standard statistical learning paradigm, researchers first expose participants to structured information—for example, shapes that appear one at a time in predictable sequences comprising a small number of items (e.g., two or three simple shapes). Later, participants are asked to decide which of two sets of items (e.g., this pair of shapes or that one) are more familiar: one that exactly matches the most reliable structure from exposure (hereafter, "old") or one in which the items have been rearranged to break some aspect of this structure ("new").

How might participants arrive at a correct choice? For example, given perfectly predictable three-shape sequences, or "triplets,"

This article was published Online First November 15, 2021.
Tess Allegra Forest https://orcid.org/0000-0002-3884-2786
A portion of these data were presented at conferences (Canadian Society for Brain, Behavior, and Cognitive Sciences, 2019; Interdisciplinary Advances in Statistical Learning, 2019; and the 42nd Annual meeting of the Cognitive Science Society) and appeared in associated conference proceedings (Proceedings of the 42nd Annual meeting of the Cognitive Science Society).

This work was funded in part by a Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant (RGPIN-2018-04933), Canada Foundation for Innovation JELF, Ontario Research Fund (36876), and University of Toronto startup funds to Margaret L. Schlichting. We would also like to thank Anna Schapiro and Noam Siegelman for their insightful discussion, as well as members of the Learning and Neural Development and Budding Minds labs. We extend particular thanks to Zahra Abolghasem for her help with data collection.

Correspondence concerning this article should be addressed to Tess Allegra Forest, Department of Psychology, University of Toronto, Saint George Campus, 100 St. George Street, 4th floor, Sidney Smith Hall, Toronto, ON M5S 3G3, Canada. Email: tess.forest@mail.utoronto.ca

participants could know the three shapes either (a) occurred in a particular, fixed order, or (b) generally occurred close together in time, with no memory for their particular order. Either sort of memory signal would favor the old sequence during a forced choice test and for this reason, correctly making such a judgment does not reveal much about the underlying representations. Do participants endorse a given sequence over another because they recognize the specific ordering of the shapes, recognize that the shapes generally occurred close together during exposure, or both?

Given the demonstrated role of the hippocampus in statistical learning (Covington et al., 2018; Finn et al., 2019; Schapiro et al., 2017; Schlichting et al., 2017), we reasoned that the memory codes observed in other hippocampus-dependent learning tasks might also be reflected in the representations of statistical structure. Drawing inspiration from episodic memory research, we therefore assessed three possibilities for the statistics extracted from a structured stream of shape triplets. The first possibility was memory for transitions, in which later memory reflects an accumulated record of specific shape-to-shape transitions (Eichenbaum, 2014; Howard & Kahana, 2002; Konkel et al., 2008; Treves & Rolls, 1992). The second was triplet membership, in which commonalities across repeated triplet presentations are emphasized at the expense of specific transitions to yield, at its most extreme, order-irrelevant memory for groups (Gilboa & Marlatte, 2017; Schlichting et al., 2015). The third was memory for position, in which shapes are "tagged" with their ordinal within-triplet position (Davachi & DuBrow, 2015; Hsieh et al., 2014) such that items occupying a given position across triplets become interchangeable in memory. We walk through each possibility in turn.

Our first candidate statistic was the particular item-item transitions embedded in the statistical stream. Learning specific item-item transitions is how statistical learning has traditionally been theorized to operate (Aslin & Newport, 2012; Aslin et al., 1998; Saffran et al., 1996; Saffran et al., 1996). Accordingly, much work in statistical learning has described it as a system that allows learners to gain sensitivity to the dips in transitional probability occurring at unit (i.e., group; here, triplet) boundaries (Aslin et al., 1998; see Thiessen, 2017 for discussion). It is important to note that such sensitivity could arise from either calculating specific probabilities, or by storing a record of all pairwise transitions and comparing new transitions to the accumulated strength of previously seen pairs (see below). Regardless, this perspective makes clear predictions about what people should find familiar on later encounter—namely, those experiences which match the transitions experienced during learning.

The second possibility was that participants might represent items as belonging together in a unit, irrespective of their particular order. This possibility is more aligned with chunk-based theories of statistical learning (Perruchet, 2019; Perruchet & Pacton, 2006; Thiessen, 2017), which suggest pairwise associations between individual items are learned *without* learning the predictive relationships between them (Endress & Mehler, 2009; Slone & Johnson, 2018). Learners may start by storing small, two-item chunks—for example, in an ABC triplet participants would first learn "A goes with B"—and gradually incorporate new items into those existing groups to learn the full pattern (i.e., "AB goes with C"; Perruchet & Vinter, 1998). Although this perspective has not made explicit claims about whether knowledge about the *order* of individual items is maintained (Perruchet, 2019), order is always held constant in the computational models used to assess chunking. As such, the underlying assumption is that order is maintained, and thus the work that has been carried out as a test of this chunking theory cannot address a key aspect of the current question. Specifically, we are interested in group knowledge above and beyond specific ordering.

In addition, while representation of unordered group information has been observed at the neural level in statistical learning tasks (Schapiro et al., 2013, 2016), it has been found only in cases where the input is intentionally devoid of other structure—that is, when reliable item-item transition information is absent. Thus, it is unclear whether unordered groupings (a) arise when typical within-triplet predictable structure is present and (b) can be detected behaviorally. Importantly, certain models of memory emphasize the importance of learning reliably co-occurring groups of items independent of their order. For example, group representation might emerge from mechanisms like those proposed for event segmentation (DuBrow & Davachi, 2013; Zacks & Swallow, 2007), or from the similarity of temporal context (Howard & Kahana, 2002), as items from the same unit occur closer together in time than do items from different units, across the entirety of the statistical stream. Neural representations in memory regions (here, the hippocampus) illustrate that we can extract regularities like category structure (Bowman & Zeithamova, 2018; Mack et al., 2016), or predictable groupings (Schapiro et al., 2017) revealed across many experiences; however, this type of learning may come with a loss of specific experiential details (Gilboa & Marlatte, 2017). Thus, if this sort of representation is stored following statistical learning, memory should reflect which items appeared in reliable groups, regardless of whether the specific item-item transitions are held intact.

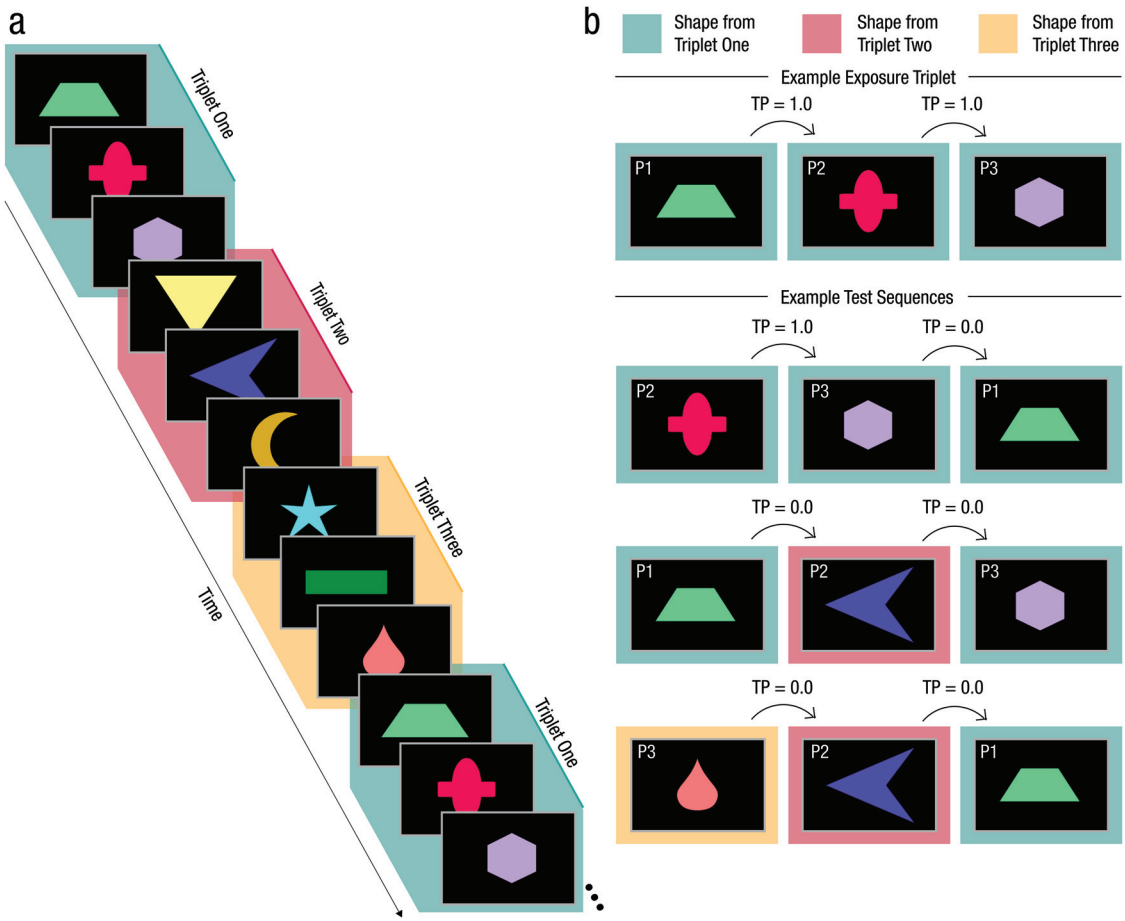
Finally, we asked whether participants store information about the ordinal position of particular shapes within their triplet. We investigated this given the centrality of temporal information in organizing our memories (Baldassano et al., 2017; Eichenbaum, 2014; Kurby & Zacks, 2008). For example, when ordinal position in a sequence is explicitly trained, eye movements during a later memory test show evidence for position coding (Pathman & Ghetti, 2015). There is also neural evidence that in addition to sequence identity, position within a fixed sequence is represented in sequence learning tasks (Hsieh et al., 2014; Kikumoto & Mayr, 2018). These findings suggest that position may also influence how we represent our experiences following statistical learning.

In addition to understanding which of these factors contribute to memory representations following statistical learning, an open question is whether memory for each factor might emerge after different amounts of structured experience. While evidence of learning has been observed after very brief experience (Batterink et al., 2015), these markers of learning were general in that they reflected only whether participants expected the next predictable item; for this reason, these data do not speak to more nuanced aspects of the representation formed by learners. We reasoned that transition knowledge might precede group knowledge, either because exclusively item-item transitions are represented in memory from the earliest demonstration of learning (Aslin et al., 1998), or because participants have learned two-element chunks, but not yet amalgamated those with other items (Giroux & Rey, 2009; Perruchet, 2019). Early sensitivity to item-item pairs is also consistent with other models of memory which predict rapid formation of memories for the specifics of experiences over which we can later

generalize (McClelland et al., 1995). This framework suggests order-independent group knowledge might come to shape memory representations only *after* specific item-item associations. Although there are no clear theoretical predictions about when position information might emerge relative to specific or group-level information, we reasoned that significant experience may be needed to first segment the stream into units before a learner can learn within-unit positions—thus, position might be particularly late-emerging.

In order to assess the independent contributions of our three candidates—item-item transitions, order-irrelevant group information, and position—on old-new judgments we systematically manipulated the degree to which probe sequences during a memory test matched participants' prior experience along these dimensions. We first had young adult participants (total N = 300) passively view a series of shapes that were, unbeknownst to them, grouped into three-shape units (triplets) that always appeared together in a fixed order (Figure 1a). To ask how the representations are influenced by experience, we manipulated the duration of initial exposure across participants (N = 100 in each of three conditions) to include either six (short exposure condition), 30 (medium), or 90 (long) repetitions of each triplet. Then, all participants made a series of old-new judgments about test sequences designed to allow us to quantify the respective impact of each of the three dimensions as well as amount of exposure on behavior

Figure 1
Task Design



Note. (A) Exposure phase; (B) example test sequences. Three three-shape triplets appeared one at a time in a pseudo-random order 6 (short exposure), 30 (medium), or 90 (long) times each. Colored backgrounds are meant to highlight triplet structure and are for display purposes only.Test sequences varied along three dimensions: Total transition score of the test sequence, defined as the sum of the transitional probabilities ("TP" in figure) of the first and second transitions in the sequence (denoted with arrows in the figure), number of shapes from the same exposure triplet (triplet membership; colored backgrounds), and number of items that maintained their within-triplet position relative to exposure (position score; white inset text). The top row depicts an "old" triplet from exposure. The bottom three rows depict example "new" test sequences which vary in their match with previous exposure. For display, the sequences chosen primarily match along one dimension. Top: high match in triplet membership (position score = 0, transition score = 1, triplet membership = 3); Middle: high match in position (position score = 3, transition score = 0, triplet membership = 2); Bottom: low match on all dimensions relative to average score across all test sequences (position score = 0, transition score = 0, triplet membership = 1). See the online article for the color version of this figure.

(Figure 1b). We then asked whether the likelihood that participants made an "old" response to each test sequence depended significantly on the degree to which it matched with exposure along each dimension. Our experiments were preregistered by exposure length, with any deviations from our plans noted throughout the results. Preregistrations along with all data and stimulus materials are publicly available (https://osf.io/ynqw6/).

Method

Participants

All participants were recruited and run online using Prolific (https://prolific.ac). Three 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 (Short exposure group: *M* age = 28.77, *SD* = 4.70, medium exposure group: *M* age = 28.12, *SD* = 4.68, long exposure group: *M* age = 28.5, *SD* = 4.46). An additional 89 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). An additional one hundred participants were included in our control experiment (described below, *M* age = 27.71, *SD* = 4.85; 38 additional excluded based on attention checks). All procedures were approved by the Research Ethics Board at the University of Toronto. Previous studies of statistical learning and memory (Batterink et al., 2015) have reported effect sizes in the medium to large range; as such, we used an effect size of $d = .60$ for determining our target sample size. Our power calculation showed that 66 subjects was required for 80% power in a multiple regression with three predictors. However, given our more subtle 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 (in each of three exposure durations) to ensure enough participants to avoid type I error.

Stimuli

All stimuli were presented to participants on their own computers using Inquisit 5.11 (2018, https://millisecond.com). Stimuli were nine shapes, each of a unique bright color. They were presented one at a time in the center 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 (Figure 1a). 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 et al., 2002).

Task Design

Exposure

Participants (N = 100 per condition) watched the stream of triplets for a total of 1.65 (short exposure condition, six repetitions/triplet), 4.95 (medium, 30 repetitions/triplet), or 14.85 (long, 90 repetitions/triplet) minutes. Participants were instructed to "*Please simply watch the images*." Shapes were presented for 1000 ms, with 100 ms interstimulus interval [ISI]. 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 (similar to, e.g., Fiser & Aslin, 2002; Saffran et al., 1996). The transitional probability (TP) between shapes 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 themselves were held constant across all participants, but the order in which participants saw the triplets varied. There were three possible orderings per exposure duration.

As this experiment was run remotely, we included multiple attention check trials (three for short and regular exposure durations, 18 for the long exposure duration) that allowed us to identify participants who were not complying with task instructions. 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. The number of attention checks was originally chosen for the regular exposure duration condition to ensure that we could monitor participants' behavior throughout the study and exclude participants even if they were slow to respond on the first attention check trial (because they could not find the "F" key, for example. The same logic was applied in choosing number of attention checks in the short exposure duration condition. The long exposure condition had a roughly proportional number of attention checks to the short exposure duration (∼1/50 shapes). Participants who failed more than one third of the total attention checks were excluded from the final sample and a new participant was run in their stead. These trials were included purely for the purpose 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.

Control Group Exposure

One could hypothesize that the structure of our test sequences themselves (described below) could lead to a particular response profile independent of any previous exposure (i.e., certain sequences of shapes could feel familiar to a particular participant even prior to exposure). Thus, we performed a separate control experiment in which 100 new participants (also recruited from Prolific using the same screening criteria as our 300 experimental subjects; see "Participants") watched a very short, scrambled version of the same shape stream prior to completing the same test as all other participants. Because there was no structure available to be learned during this random exposure, any subsequent memory test behavior would have to be driven by the structure of the test phase itself. These participants watched only six presentations of each shape—mirroring the duration of exposure experienced by our short exposure group—for two reasons. Although we saw no evidence that our short exposure duration group was more prone to changes in their representation over the course of the test (see online supplemental materials, "Confirming Main Effects Are Present Throughout Test Phase"), we still reasoned that if any test-phase representations of transition score, triplet membership, or position

were formed, they would likely have the greatest impact among the short exposure group because it was participants in this condition that had the least experience with those statistics during exposure. Thus, we wanted to ensure the experience of our short exposure and control groups were equated as much as possible. Second, our goal was to have as little exposure to any of the shapes as possible, while still allowing us to use the same test instructions ("Do you think this set of shapes is part of what you saw before?")—instructions that would not be sensible without some previous exposure phase.

Test Phase

After exposure, participants in all conditions 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 not (new). Three of the sequences were the reliable triplets (for example, ABC from exposure; hereafter termed exposure triplets), two were foils which *were* presented during the exposure stream but not as reliably as a triplet (for example, BCD; these sequences are technically old but treated as new for our purposes; see below), and 17 were new (detailed below). Because all shapes were seen at exposure, participants' judgments about test sequences could not be based on familiarity with the shapes themselves, but rather had to be made based on the particular sequence of shapes. All test sequences were presented with each item appearing alone for 1000 ms with a 100 ms ISI to match 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. Each participant saw a different order of test trials.

Test Sequences

The goal of this experiment was to understand which 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: (a) the sum of the TP of each forward transition (TP from the first item to the second, and the second to the third) in a test sequence ("transition score"); (b) the number of shapes in a test sequence that had been part of the same exposure triplet ("triplet membership"), and (c) the total number of shapes which maintained their position within a test sequence relative to their exposure triplet (first, second, third; "position"). The specific sequences were generated so as to orthogonalize the relative contributions of these three dimensions as much as possible. 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; Figure 1b). 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 (e.g., BAC [2–1-3] or IGH [3–1–2]); six were sequences with two shapes from one exposure triplet (e.g., AEC [1–2–3] or IDE [3–1–2], a subset of these were the part-word foils mentioned above); and six were sequences with one shape from each exposure triplet, akin to the "position-matched foils" commonly used in visual statistical learning paradigms (e.g., Park et al., 2018; Schlichting et al., 2017; Turk-Browne et al., 2005; e.g., AEI [1–2–3]; these could also be position-matched to a shuffled sequence: IGB [3–1–2]).

Finally, we calculated the total transition score for each sequence. For all sequences with a triplet membership score of 1 or 3, transition score 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 transition score) for the test sequences with a triplet membership score of 2, we split these sequences into two groups with different transition score 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 transition score of 1.5. The other four position orderings formed group two—here, the two same-triplet shapes selected created a test sequence with a transition score of 0 in order to most closely match their position ordering counterparts with triplet membership scores of 1 and 3 (e.g., AEC).

As noted above, some of our critical test sequences (specifically "part-word foils") were in fact part of what participants had seen during exposure, and thus an "old" response is not technically incorrect for these sequences. That said, our analysis approach (described below) differs from previous work in that we manipulated each factor parametrically to maximize our ability to isolate the separate effects of transition score, triplet membership, position. Thus, we are focused on understanding how subtle increases in one dimension relate to old judgments rather than scoring particular test responses as correct or incorrect. As such, we consider part-word sequences and other truly new sequences together for analysis purposes as they were not the reliable groupings that participants had seen. We will refer to this set of test sequences (part-word sequences and all new sequences) as "foils" from here on. In addition, we have ensured that our effects are not driven by simply a tendency of participants to make "old" responses to part-word foils (online supplemental materials, "Do Memory Judgements Track With Binary 'Seen' Status?").

Additionally, while our experimental design was tailored to assess *forward* transitions (the probability that A is followed by B), we also consider bidirectional associations (given exposure to

"AB," whether "BA" is endorsed as old at test) in follow-up analyses in the online supplemental materials (online supplemental materials, "Bidirectional Item-Item Links Do Not Predict Responses Better Than Forward Transitions"). Note that the bidirectional associations we interrogate here (as in Park et al., 2018) are in contrast to the more frequently investigated—and similarly named, but distinct—concept of *backward* transitions (the probability that B was preceded by A; see e.g., Endress et al., 2020; Tummeltshammer et al., 2017), which our study is not poised to measure.

Data Analysis

Considering each factor in isolation revealed that transition score, triplet membership, and position were all positively related to the tendency to endorse a test sequence as old (mixed effects models, collapsed across exposure duration, which predicted old responses as a function of just one factor; old ∼ transition score: $\beta = .66$, $SE = .05$, $z = 14.05$, $p < .0005$; old ∼ triplet membership: $\beta = .47$, $SE = .03$, $z = 17.07$, $p < .0005$; old ∼ position: $\beta = .03$, $SE = .006$, $z = 5.13$, $p < .0005$). However, these simple models would not have captured the independent contributions of each factor, as there are some unavoidable dependencies between factors (e.g., when position is high, transition score must be low). Thus, we chose our analysis approach to isolate the respective contributions of transition score, triplet membership, and position to old-new judgments at test, after statistically controlling for the other factors.

Specifically, we constructed two preregistered generalized mixed-effects models (logistic regression) using the glmer function in the lme4 package (Bates et al., 2015; R Core Team, 2020) in R Version 3.5.1, to investigate the degree to which each factor was associated with the probability of making an old response (see "Modeling Transition Score, Triplet Membership, and Position"). Because we were also interested in how these factors came to influence memory after differing exposure durations, we also included exposure duration as an unordered factor in the model and ran smaller models on each exposure duration separately (see "Effects of Exposure Duration").

Modeling Transition Score, Triplet Membership, and Position

To address our overarching question of which factor(s) contribute to memory after controlling for the others, we ran two models at different levels of complexity to ask which better accounted for the data: first, a main effects model and second, a model that included a subset of all possible two-way interaction terms. Both models included fixed effects for the main effects of transition score, triplet membership, and position as continuous numeric predictors, as well as exposure duration as an unordered factor. The interactions between exposure duration and each factor of interest were also included. The second model additionally included interactions between triplet membership and position and between triplet membership and transition score (see below). In both models, random by-participant slopes and intercepts were included for each of the within-subjects terms (including Transition Score × Triplet Membership and Position × Triplet Membership interactions). In both models, the inclusion of exposure duration represents a slight deviation from our preregistered analysis plan, as exposure duration was identified as a variable of interest after we had already completed one experiment, and thus each exposure duration was preregistered independently; otherwise, our model setup and comparison approach are identical to our preregistration.

The interaction between triplet membership and transition score was included in the second model in order to account for the possibility that the degree to which transition score informed old-new decisions would depend on the number of items from an exposure triplet present in the test sequence (triplet membership). For example, transitions between the individual shapes might matter more for test sequences which had a triplet membership score of 0—for which transitions are 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 of 0, intact transitions might be especially likely to underlie any erroneous "old" responses. Similarly, the interaction between triplet membership and position was included in the second model to account for the possibility that position might be more important in test sequences with higher triplet membership. 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 0, in which none of the shapes had occurred in the same triplet during exposure.

We compared the model with only main effects to the model 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 ($BIC_{\text{Main Effects Model}} = 6644.3$, $BIC_{\text{Interaction Model}} = 6627.7$, $\chi^2[4] = 24.55$, $p < .001$), so we focus here on the results of this better-fitting model. However, the directionality and overall significance of all three main effects was similar between the two models. Results are reported using the Anova function from the Car package (Fox & Weisberg, 2019; reference level = medium exposure duration).

These models were run excluding truly old, exposure triplet sequences because we were interested in how subtle parametric increases along each of our dimensions of interest impacted the likelihood endorsing an item as "old," rather than broad categorical differences in responses between the exposure triplets and foil test sequences described above. Moreover, because exposure triplets represent the highest possible value on each of our metrics, responses to these sequences do not help adjudicate between which signal is driving old-new memory decisions. Our choice to exclude exposure triplets from the analyses means that any linear relationships observed cannot be disproportionately driven by responses to these truly old sequences, nor by major differences in how participants respond to truly old sequences as compared to foils. This differs from the preregistered analysis plan, but results were similar when we also included old sequences in the analyses as we had preregistered (reported in online supplemental materials, "Results do not change when exposure triplets are included in analysis").

Effects Within Each Exposure Duration

To address our second question of whether representation of each of these factors emerges simultaneously, we additionally ran our model on the data from each exposure duration separately, following our preregistered analysis plan for each exposure duration.

These results are reported alongside the direct comparisons between exposure durations, which was not part of our preregistration.

Reporting Effect Sizes

For all beta values in this study, results were 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.

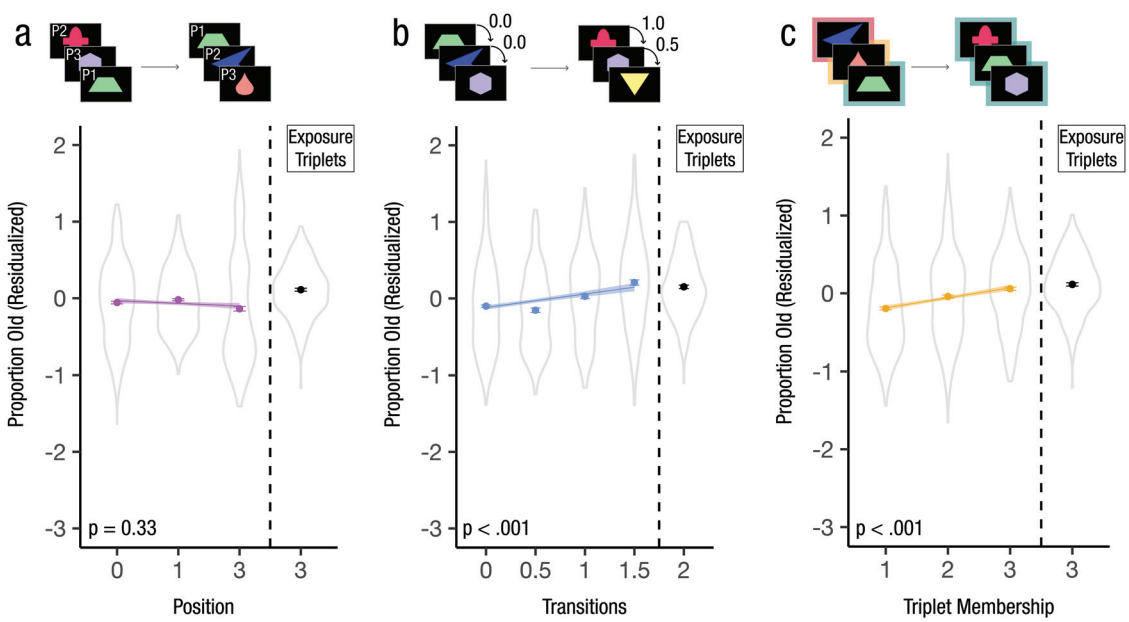
Results

We first confirmed that participants were able to learn the statistical structure by verifying that across all exposure durations, exposure triplets were endorsed as old more frequently than all other sequences during the memory test (mean proportion old for exposure triplets: M_{old} = .66, mean proportion old for all other test sequences: M_{new}= .43, $t[299] = 12.28$, $p < .001$, $d = 1.07$). Memory performance was also reliably above chance at each of the three different exposure durations separately (short: M_{old} = .62, M_{new} = .47, $t(99) = 6.53$, $p < .001$, $d = .85$; medium: M_{old} = .68, M_{new} = .41, $t(99) = 8.33$, $p < .001$, $d = 1.25$; long: M_{old} = .67, M_{new} = .44, $t(99) = 6.59$, $p < .001$, $d = 1.09$). Given this confirmation that participants had learned the exposure triplets, we next asked how our factors of interest—transition score, triplet membership, and position—might explain variability in behavioral responses among the foil sequences. In particular, the subsequent analyses asked whether a match along any of these dimensions might be associated with an elevated tendency to endorse a foil sequence as "old."

Both Specific Transitions and General Groupings Contribute to Memory Judgments

One of our central questions was whether transition score, triplet membership, and position would impact participants' judgments of whether a foil test sequence was old, irrespective of exposure duration. Results showed that across all exposure durations, position had no effect on the likelihood of a test sequence being endorsed as old ($\chi^2_1 = .95$, $p = .33$, Figure 2a). In other words, maintaining specific items in their position from exposure had no measurable effect on memory judgments when controlling for the other two factors. However, both transition score and triplet membership did have reliable positive effects on participants' old-new judgments (transition score: $\chi^2_1 = 23.20$, $p < .001$, Figure 2b; triplet membership: $\chi^2_1 = 20.66$, $p < .001$, Figure 2c): Participants were more likely to make old responses as a function of the number of both 1) transitions held intact and 2) shapes that

Figure 2
Transition and Triplet Membership Scores Track Old-New Judgments Across All Exposure Durations



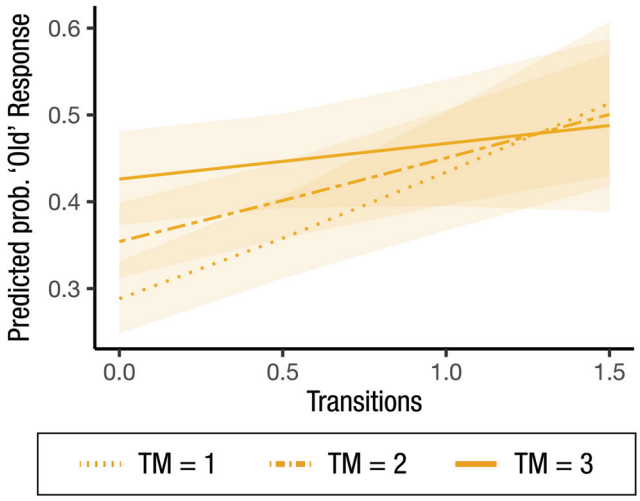
Note. Memory judgments as a function of (a) position (purple), (b) transitions (blue), and (c) triplet membership (yellow) across all exposure durations. There were reliable effects of transition score and triplet membership on memory, but position did not impact memory. For all panels, example test sequences at minimum and maximum values of each dimension are provided on top of the graphs to show the range of possible test sequences on each dimension (excluding truly old sequences). Y-axis represents residualized proportion old responses after regression on the other two factors. Gray violins represent distribution of participant responses at each level of the predictor dimensions; dots show mean values along with 95% confidence intervals. Lines 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. Shaded area around lines also represents 95% confidence intervals. See the online article for the color version of this figure.

occurred in a triplet together at exposure. These results are in line with the idea that we form both specific and general representations of the statistical structures we experience, even after explicitly controlling for the alternate type of representation.

As mentioned, we also assessed whether transition score or position impacted old-new decisions more when test sequences contained more items from the same group across all exposure durations. The interaction between triplet membership and position was not significant ($\chi^2_1 = 1.44$, $p = .23$), indicating that the effect of position did *not* depend on the number of items in a test sequence that maintained their position relative to exposure. There was, however, a significant interaction between transition score and triplet membership ($\chi^2_1 = 5.32$, $p = .02$), such that there was a larger effect of triplet membership when transition score was low, and conversely, a larger effect of transition score when triplet membership was low (see Figure 3). The finding that triplet membership is more influential when relatively fewer transitions are held intact (and vice versa) might suggest that when there is minimal signal from either source, participants rely more heavily on a combined signal of both transition score and triplet membership.

Overall, these results suggest that while position may not impact old-new judgments transition score and triplet membership each make reliable contributions to memory behavior on a trial-by-trial basis within participant. This adds to our understanding of what we learn from structured experience by suggesting that both specific item-item relationships and order-irrelevant group information come to be represented in memory following statistical learning. Moreover, this order-irrelevant group information may be particularly useful for guiding memory decisions when specific item-item relationships are not available.

Figure 3
Transition Score and Triplet Membership Interact to Predict Old Responses



Note. Predicted probability of responding old to a particular test sequence (Y-axis) as a function of transition score (X-axis) and total triplet membership (abbreviated as TM; line style) across all exposure durations. There was a larger effect of triplet membership when transition score was low, and vice versa. Shaded bands around the lines represent 95% confidence intervals. See the online article for the color version of this figure.

When in Learning Do These Representations Emerge?

Understanding the order in which these representations emerge—or whether they might be formed in parallel—was also of central interest. Statistical learning tasks can be ended at any moment with limited impact to the structure of the experiment overall and no change to the number of test trials. This makes it a particularly elegant test case for understanding how the nature of learning depends on the amount of exposure. In fact, the number of times participants are exposed to each unit has varied widely in past research, from around 24 in many visual statistical learning studies (e.g., Park et al., 2018) to as many as 2,240 in the auditory modality (Finn & Hudson Kam, 2008). Successful statistical learning has been demonstrated after as few as three exposures to each unit (Batterink, 2017), but no work has been done to understand how exposure duration changes the representations formed during learning in adults (although see Slone & Johnson, 2018 for similar questions addressed in infants). Thus, by systematically stopping exposure after only a limited number of repetitions of each triplet (here, six), we can shed light on whether knowledge of individual pairwise relationships precedes triplet membership knowledge, as may be predicted by early learning of two-item chunks or early knowledge of item-item transitions. We also included a long exposure condition to test the idea that extensive experience (operationalized here as 90 presentations of each triplet; though note that this is within the range of exposure durations frequently reported in other research) may be needed to represent position in particular, which inherently requires participants to have learned which items form a unit (e.g., by showing memory for triplet membership) before they can assign particular positions to items within those units.

To investigate these possibilities, we now report interactions between exposure duration and any of our factors of interest (from the same model described above). We also report the effects of transition score, triplet membership, and position on the data from each exposure duration separately.

Main Effects of Exposure Duration

There was a main effect of exposure duration on the overall likelihood of making an old response to new items ($\chi^2_1 = 7.64$, $p = .02$): Participants in the short exposure condition were more likely say "old" than those in the other two conditions (short vs. medium exposure: $\beta = .43$, $SE = .16$, $z = 2.68$, $p = .007$, Odds Ratio = 1:1.53; short vs. long exposure: $\beta = -.30$, $SE = .16$, $z = -1.91$, $p = .06$, Odds Ratio = 1:.74; medium vs. long exposure: $\beta = .13$, $SE = .16$, $z = .78$, $p = .44$, Odds Ratio = 1:1.1), suggesting that less exposure leads to more memory errors to foil test sequences overall.

Specific Transition Knowledge Emerges With Experience and Is Maintained

Both transition-based and chunking accounts of statistical learning might predict that memory for the one-to-one relationships between items emerge early in learning, either because participants have learned some transitional information or because they have learned about smaller two-item chunks within the triplet. Contrary to this idea that item-item transitions guide memory from early in learning, we found that the influence of transition score on memory varied across exposure duration (transition score × exposure

duration interaction, $\chi^2_2 = 9.78$, $p = .007$; Figure 4). In particular, the effect of transition score was significantly smaller at the short exposure duration as compared with both medium ($\beta = -.34$, $SE = .12$, $z = -2.76$, $p = .006$, Odds Ratio = 1:.71) and long ($\beta = .32$, $SE = .12$, $z = 2.59$, $p = .01$, Odds Ratio = 1:1.38); medium and long exposure duration were not reliably different from one another ($\beta = -.02$, $SE = .13$, $z = -.19$, $p = .85$, Odds Ratio = 1:.98). Investigating each exposure duration separately, there was no effect of transition score on old-new judgments after short exposure ($\beta = .32$, $SE = .18$, $z = 1.79$, $p = .07$, Odds Ratio = 1:1.38); however, transition score did significantly predict old-new judgments after both medium and long exposure (medium: $\beta = .70$, $SE = .20$, $z = 3.58$, $p < .001$, Odds Ratio = 1:2.02; long: $\beta = .60$, $SE = .21$, $z = 2.90$, $p = .004$, Odds Ratio = 1:1.82).

Collectively, these results suggest that once knowledge of item-item transitions emerges sometime after our shortest exposure duration, transition score remains important for guiding memory judgments after both medium and long exposure. Interestingly, this means that while participants demonstrate learning after minimal exposure (as reported earlier, $M_{old} = .62$, $M_{new} = .47$, $p < .001$, & in line with previous research Batterink, 2017), this behavior after early learning may not track with the total transition score in a test sequence.

Early Emergence of Group Knowledge

Given that participants are able to perform above chance and yet transitions are not driving old-new judgments after short exposure—what is? One possibility is that a fuzzy representation of simple co-occurrences might be formed early on and support a general sense of oldness, as has been suggested for other memory tasks (Ahmad et al., 2017; Wilburn & Feeney, 2008). Supporting this, there was no significant interaction between triplet membership and exposure duration ($\chi^2_2 = .20$, $p = .90$, Figure 4); rather, triplet membership was a significant predictor of old-new judgments at every exposure duration (short: $\beta = .43$, $SE = .12$, $z = 3.58$, $p < .001$, Odds Ratio = 1:1.54; medium: $\beta = .33$, $SE = .13$, $z = 2.51$, $p = .01$, Odds Ratio = 1:1.39; long: $\beta = .31$, $SE = .12$, $z = 2.52$, $p = .01$, Odds Ratio = 1:1.36).

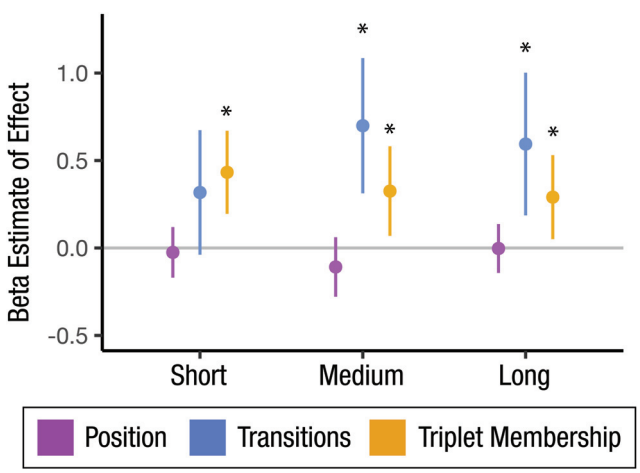
These results suggest that even from six presentations of each triplet, triplet membership already informs old-new decisions. In contrast to memory for transitions—which increases initially and is then maintained—memory for triplet membership seems to be present from even the earliest point in learning and remains stable over exposure. Notably, because memory for triplet membership emerges before transition knowledge, these results also imply that representing triplets as a cohesive unit does not require memory for the underlying item-item transitions.

To better understand whether these early representations arise from a general sense of oldness that is independent of ordered item-item transition knowledge, we ran an additional exploratory mixed effects model (full model is detailed in the online supplemental materials, "Model Specification for Shuffled vs. Triplet Sequences Model") including only exposure triplets (e.g., ABC, which were removed for the main mixed model analyses) and shuffled test sequences (e.g., BCA). If there is an initial step in which order is fully irrelevant, these two test types should be treated as equally old initially and come to be treated as increasingly different with extended exposure. This model revealed that there was a main effect of test type such that exposure triplets were treated as older than shuffled sequences ($\chi^2_1 = 48.75$, $p < .001$), suggesting that early representations may still include *some* information about the relative order and/or position information for within-group items. There was, however, also an interaction between test type and exposure duration (test type × exposure duration, $\chi^2_2 = 6.11$, $p = .046$), such that these two test types were treated more differently after medium than short exposure (all other pairwise exposure duration comparisons were nonsignificant, all p's $> .12$). This interaction between exposure duration and test type provides some additional support for the idea that early representations are *more* independent of ordered item-item transition knowledge than memory representations following extended exposure. We expand upon this further in the Discussion.

No Effect of Position at Any Exposure Duration

While there was no main effect of position, it is possible that position would only begin to impact old-new judgments after participants had learned the triplet membership structure. However, there was no interaction between position and exposure duration ($\chi^2_2 = .68$, $p = .72$, Figure 4). Moreover, position was not a significant predictor of old-new judgments at any exposure duration (short: $\beta = -.02$, $SE = .07$, $z = -.33$, $p = .75$, Odds Ratio = 1:.98; medium: $\beta = -.11$, $SE = .09$, $z = 2.51$, $p = .01$, Odds Ratio = 1:.90; long: $\beta < .001$, $SE = .07$, $z = .006$, $p = .99$, Odds Ratio = 1:1.00), meaning that even after 90 presentations of each triplet, we found no evidence that position score was being used for old-new decisions.

Figure 4
Exposure Duration Impacts Memory



Note. Beta estimates (slopes from mixed effects models, as shown in Figure 2) of position (purple), transitions (blue), and triplet membership (yellow) for each exposure duration. Our transitions metric was the only factor to show a significant interaction with exposure duration, such that it was reliably different from chance only in the medium and long exposure duration conditions. Triplet membership was significant at all exposure durations, in contrast to position which was not significant at any exposure duration. Beta estimates (dots) and 95% confidence intervals (error bars) were derived from models run on each exposure duration separately. All asterisks represent p-values less than .05. See the online article for the color version of this figure.

No Evidence of Learning During Test After Structured Exposure

It is possible that test-phase experience after structured exposure might change the statistics learners use to make old-new judgments after statistical learning. As such, we interrogated these data for possible changes in response patterns as participants gained more test-phase experience, in order to examine whether participants were learning *any* structure that may have been present during test. Although test sequences were presented in a different random order for each participant, each third of the test phase contained one presentation of each test sequence. This means we could model the effects of transition score, triplet membership, and position equally well in each test-third. Thus, we added test-third as a predictor to our model described above (online supplemental materials, "Confirming Main Effects Are Present Throughout Test Phase"), along with the interaction terms between test-third and each factor of interest to ask whether there are differences in the relationship between transition score, triplet membership, or position as a function of test experience.

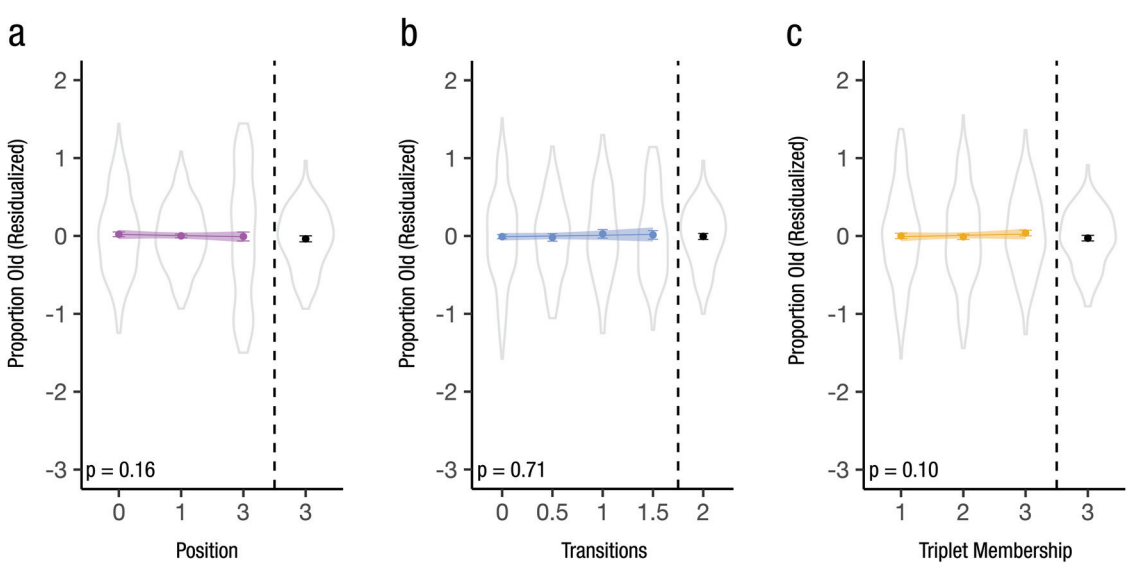
Collapsed across all exposure durations, there was no significant interaction between test-third and any factor of interest (all χ^2 values < 2.1, all p's $> .15$), indicating that the relationships between each of our factors and memory decisions did not significantly change as participants gained more experience with the test sequences. Moreover, main effects (collapsed across exposure lengths) were similar to the original analysis across the entire test, such that there were reliable effects of transition score ($\chi^2_1 = 4.40$, $p = .04$) and triplet membership ($\chi^2_1 = 6.54$, $p = .01$), but not position ($\chi^2_1 = 2.06$, $p = .15$), when controlling for effects of test-third. Additional analyses examining the interaction of each effect of interest and test-third in each exposure duration separately, as well as after each test-third separately (both analyses found in online supplemental materials, "Confirming Main Effects Are Present Throughout Test Phase") also indicate that our results were present from the beginning of the test phase. As such, test phase experience does not appear to impact the reported patterns of data.

Test Experience Is Not Sufficient for Representations to Emerge

While our analyses of test-third suggest that participants' responses are not changing over the course of test, we wanted to ensure that the memory representations participants used to guide their responses were truly formed during exposure to structured information, and not during the test phase. Thus, we ran an additional control experiment (as noted in the Methods), in which participants watched a very short, scrambled version of the same shape stream prior to completing the same test as all other participants. Because there was no structure available to be learned during this random exposure, any subsequent memory test behavior would have to be driven by the structure of the test phase itself.

Results of this control experiment revealed that unlike any of our structured exposure groups, the random exposure group did not differ in how likely they were to call foil and exposure triplet sequences old ($M_{old} = .48$, $M_{new} = .51$, $t[99] = -1.62$, $p = .11$, $d = .16$). Importantly, there were no effects of transition score, triplet membership, or position after random exposure (all $\beta < abs[.19]$, all $p's > .10$, Figure 5). This underscores that in the main experiment, even after the shortest exposure duration, the responses we observed were a function of the learning phase itself, rather than based on memory representations formed solely during test.

Figure 5
Structured Exposure is Required for Effects to Emerge



Note. No effect of position, transitions, or triplet membership in the absence of structured experience. Results of control experiment revealed no reliable effects of position (a), transitions (b), or triplet membership (c) following a randomly ordered six presentations of each shape. Data are presented as in Figure 2. See the online article for the color version of this figure.

Discussion

We show that memory judgments after statistical learning reflect both item-item transition information *and* order-irrelevant groupings, but not position. To our knowledge, this is the first behavioral demonstration that order-irrelevant group information influences memory beyond item-item transitions. Additionally, we show that group knowledge comes to shape memory judgments quite early in learning—sometime between one and six exposures to each item—while parametric increases in the consistency of item-item transitions (relative to exposure) take between six and 30 exposures to shape memory. Even after 90 exposures to each item, position did not impact memory judgments.

In line with previous research suggesting the importance of temporal item-item links for statistical learning (Aslin et al., 1998), we found that test sequences with high forward transition scores were more likely to be endorsed as old. While we operationalized transition scores as the sum of forward TPs, similar analyses using bidirectional associations can also be found in the online supplemental materials ("Bidirectional Item-Item Links Do Not Predict Responses Better Than Forward Transitions"). This finding converges with the notion of temporal associations 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 et al., 2016). We show old-new memory decisions reflect these links, building on the past statistical learning literature which has largely measured *relative* oldness using forced choice tests (cf. Bays et al., 2016; Isbilen et al., 2017; Park et al., 2018; Turk-Browne et al., 2005). This departure from standard testing means we are now able to show that people are sensitive to even small variations in the extent to which test sequences match their prior experience.

Old-new decisions also tracked with the number of items from the same group when controlling for other factors, highlighting another core memory function at play during statistical learning: grouping items into meaningful units across our experiences. We conclude that even when item-item transitions are a useful cue for learning, representations formed as a result of statistical learning go beyond a set of item-item links to include order-irrelevant group information. 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 et al., 2012)—even if such co-occurrences are indirect (Luo & Zhao, 2018; Schlichting et al., 2015). Moreover, once we have linked these co-occurring elements, our associative memory for items within one event is stronger than memory for items that span event boundaries (Ezzyat & Davachi, 2011), in the same way items in one group were more strongly linked with each other in our data. We propose a similar mechanism is at play in statistical learning (Schlichting et al., 2017), yielding a general triplet representation that emphasizes general group membership to influence behavior.

Our finding that group membership is stored in memory is also in line with other sequence learning work suggesting that transitions, group membership, or position can be represented in memory if it is the *only* type of structure available to learn. For example, when the strength of item-item transitions are uninformative, learners represent group information (Schapiro et al., 2013, 2016); moreover, even subtle differences in the order in which this information is presented can bias the representations formed (Karuza et al., 2017; a similar phenomenon has also been observed for position [Kikumoto & Mayr, 2018]). These results are all consistent with the notion that representations flexibly adapt to reflect the type of information available. Here, we illustrate that in situations where multiple statistics *could* guide memory, learners use multiple types of representations to make their memory decisions.

The weight carried by each signal may moreover adapt according to the strength of the other signals present. Specifically, we saw an interaction between triplet membership and transition score (see Figure 3*)*, with triplet membership having greater influence on memory decisions when transition score was low and vice versa. There are at least two possible interpretations of that finding, both of which somewhat align with our results: (a) Triplet membership may be a secondary signal to specific item-item associations, perhaps reflected in the relatively greater effect for transition score than triplet membership; or (b) specific item-item associations may be a secondary signal to triplet membership, consistent with the earlier emergence of sensitivity to triplet membership as compared with transition score. Whether this interaction has a particular directionality—or rather, whether both types of signals can compensate when the other is low—remains an open question for future investigation.

Our findings also add to an ongoing discussion of the extent to which statistical learning proceeds via learning transitions between individual items or "chunks" of information, which are then aggregated to learn about the larger units (Endress & Mehler, 2009; Giroux & Rey, 2009; Perruchet & Vinter, 1998). Our data provide support for the idea that item-item transitions are stored in memory. In our study, learners could either have computed the exact transitional probability between two shapes to make these judgments, or they could have stored a record of all experienced pairwise transitions[1] against which they could query a new pairwise transition at test, a possibility which is discussed further in the online supplemental materials, "Do Memory Judgements Track With Binary 'Seen' Status?"). Regardless, we find that there is also distortion of the input in memory such that items that co-occur (i.e., those in the same triplet) become more interchangeable than items that do not (those from different triplets)—reflected here in our measure of triplet membership. Importantly, this result was robust to changes in our specific analytic approach throughout our control analyses (online supplemental materials, "Do Memory Judgements Track With Binary 'Seen' Status?"). Our results are therefore consistent with a chunking model (Giroux & Rey, 2009), but further suggest that rather than learning specific, ordered groups of items, learners represent order-irrelevant, higher-level information, and they do so from very early in learning.

Our results mirror recent findings showing that participants endorse shuffled pairs of items (i.e., "BA" after consistently seeing

[1] While divergent from computational models of statistical learning (c.f. iMinerva by Thiessen & Pavlik, 2013), such a perspective would be in line with the many exemplar models of memory and category learning (Annis & Palmeri, 2019; Hintzman, 1984; Kruschke, 1992; Love et al., 2004; Nosofsky, 1987; Shiffrin & Steyvers, 1997; Zaki et al., 2003) that posit storage of all experiences, optionally in combination with some central tendency, and which have been interrogated using neuroimaging methods in recent years (Bowman et al., 2020; Bowman & Zeithamova, 2018; Davis et al., 2012a, 2012b, 2017; Mack et al., 2013, 2016, 2020; O'Bryan et al., 2018).

"AB," as we assessed in our bidirectional association analysis) as older than position-matched foils (Park et al., 2018); here, we extend that work by showing that maintaining more items from one group increases the likelihood of an old response—independent of order, the number of intervening items, or the comparison test item. Broadly, our conclusion that multiple signals are stored in memory following statistical learning also nicely complements recent modeling work about the computations that are used during the learning process itself, which suggests different computational strategies may be used by different people (see Siegelman et al., 2019 for further discussion).

We saw no evidence that position influences behavior after controlling for other factors. While this finding might seem at odds with prior work showing position coding in memory for sequences (Hsieh et al., 2014; Kalm & Norris, 2017), one possible reason for these apparent differences is that here, we controlled for factors that have frequently been correlated with position in past research. Moreover, one important nuance of our task is that position is initially ambiguous and becomes evident only after stream segmentation—not always the case in prior work in which sequence boundaries are provided. In other words, triplets must first be detected before within-triplet position becomes relevant. While the long exposure condition (90 presentations of each triplet) was included to allow time for position representation to emerge after triplet groupings, it is possible that even our long exposure was simply not long enough for this to occur. Future work might provide participants with extended experience—perhaps across multiple days—to assess whether position codes eventually emerge.

Interestingly, position has often been held constant (relative to exposure) in the test foils used in previous statistical learning work, particularly in the visual statistical learning literature. In many studies, the test of learning involves participants adjudicating between a correct set of three shapes (e.g.,"ABC") and a series of shapes which match previous exposure only in their position information (e.g.,"AEI"; Turk-Browne et al., 2005; Zhao & Luo, 2017). Crucially this means that many studies have likely maximized the chance of seeing any learning effect, because people are less sensitive to position. Interestingly, auditory statistical learning work frequently uses test sequences which maintain some item-item transitions and some triplet membership information (e.g.,"BCD" when participants learned "A<u>BC</u>" and "<u>D</u>EF"; Batterink et al., 2019; Saffran et al., 1996; Toro et al., 2005). Our data suggest these sorts of foils will be more difficult to discriminate from exposure triplets than the position-matched foils described above. Perhaps unsurprisingly, auditory statistical learning performance is often worse than visual statistical learning performance (see Arnon, 2020); our findings suggest this difference may be due at least in part to differences in the structure of the test sequences used, not just modality.

Our exposure duration results also illustrate that these different representations are not all formed on the same timescale. Order-irrelevant group information is present after very limited exposure, while transition information does not shape memory judgments until participants have had more exposure (see also Tompary et al., 2020 for complementary work on rapid schema acquisition). It is important to note that after short exposure, shuffled sequences were still treated as relatively newer than exposure triplets, and thus early group knowledge may still maintain some order information (either in the form of position coding or ordered transition information, which are both held constant relative to exposure) for items in the same group. Giving participants even less experience with the structure is an important avenue for future work and may reveal truly order-independent early group knowledge. That said, in tandem with our main results showing item-item information guides old-new judgments after early exposure across all test sequences in a continuous fashion, the finding that exposure triplets are treated as older than shuffled foils means that memory behavior (after short exposure) may indeed reflect order information—but importantly, only for test sequences with three shapes from the same triplet. This might suggest that early group membership knowledge could rapidly support the learning of within-triplet transitions (which would explain why participants discriminate between shuffled foils and triplets), while between-triplet transition knowledge takes longer to emerge. Regardless, future work could attempt to fully understand whether item-item and group membership representations are indeed independent, or rather emerge hierarchically such that triplet membership representations are a prerequisite to form specific item-item links. Once present, both of these representations are maintained even with more exposure, such that specific item-to-item transitions are added to triplet membership representations as an additional cue for memory with extended exposure.

Our results did not show any difference between the medium and long exposure duration groups, in any of the three representations measured here. It is worth noting that there are (at least) two possible ways additional exposure could change participants' learning. First, overall performance could differ between each group, as measured by false alarm rate to foil sequences, slower reaction times, or additional failed attention checks. Second, memory for each factor of interest could be relevant after a different amount of experience. As for the first possibility, we saw no evidence of overall performance differences between our medium and long exposure group (see first paragraph of "Results," and online supplemental materials, "Assessing Differences Across Exposure Duration Groups in Overall Performance"). The limited differences we observe are line with past research suggesting that increased exposure does not necessarily result in better overall learning (Bulgarelli & Weiss, 2016; Finn & Hudson Kam, 2008; Gebhart et al., 2009; Slone & Johnson, 2018). With regards to the second possibility, we included the long exposure group specifically to offer the opportunity for position to emerge after groups and transitions are more established. However, it was also possible that triplet membership or transition score would have been more strongly represented following extended exposure relative to medium exposure. Neither of these possibilities were borne out in our results, but it is of course possible that even more extended exposure would be needed to reveal such differences.

An open question is whether the nature of group representation specifically is the same after limited versus extended exposure. For example, early group knowledge may reflect poor fidelity memory of our early experiences (Ahmad et al., 2017; Batterink & Paller, 2017; Gomez, 2016; Wilburn & Feeney, 2008), which is less related to someone's specific transition knowledge than the group knowledge that occurs after a greater amount of experience. This interpretation is supported by our finding that after short exposure, shuffled sequences and truly old sequences are treated relatively more similarly than they are after medium exposure. Later in learning, however, triplet membership representations may arise

from emphasizing commonalities across experiences at the expense of specific transitions (McClelland et al., 1995).

Despite unequivocal evidence that learners are sensitive to statistics in their world, until now we lacked a clear understanding of how variation in the amount of experience with these statistics eventually shapes their behavior. Here, we demonstrate that when both item-item links and group information are available to support learning, learners use both to guide their behavior, starting with group information. These results provide novel insights into past statistical learning studies and an explicit link to the memory literature, putting us in a position to move forward with studying how our varied experiences shape how we come to represent the world.

References

Ahmad, F. N., Moscovitch, M., & Hockley, W. E. (2017). Effects of varying presentation time on long-term recognition memory for scenes: Verbatim and gist representations. *Memory & Cognition*, *45*(3), 390–403. https://doi.org/10.3758/s13421-016-0672-1

Annis, J., & Palmeri, T. J. (2019). Modeling memory dynamics in visual expertise. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(9), 1599–1618. https://doi.org/10.1037/xlm0000664

Arnon, I. (2020). Do current statistical learning tasks capture stable individual differences in children? An investigation of task reliability across modality. *Behavior Research Methods*, *52*(1), 68–81. https://doi.org/10.3758/s13428-019-01205-5

Aslin, R. N., & Newport, E. L. (2012). Statistical learning: From acquiring specific items to forming general rules. *Current Directions in Psychological Science*, *21*(3), 170–176.

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science*, *9*(4), 321–324. https://doi.org/10.1111/1467-9280.00063

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. https://doi.org/10.1016/j.neuron.2017.06.041

Bates, D., Machler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*. Advance online publication. 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. https://doi.org/10.1177/0956797617698226

Batterink, L. J., & Paller, K. A. (2017). Online neural monitoring of statistical learning. *Cortex*, *90*, 31–45. https://doi.org/10.1016/j.cortex.2017.02.004

Batterink, L. J., Paller, K. A., & Reber, P. J. (2019). Understanding the neural bases of implicit and statistical learning. *Topics in Cognitive Science*, *11*(3), 482–503. https://doi.org/10.1111/tops.12420

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

Bays, B. C., Turk-Browne, N. B., & Seitz, A. R. (2016). Dissociable behavioural outcomes of visual statistical learning. *Visual Cognition*, *23*(9-10), 1072–1097. https://doi.org/10.1080/13506285.2016.1139647

Bowman, C. R., Iwashita, T., & Zeithamova, D. (2020). Tracking prototype and exemplar representations in the brain across learning. *eLife*, *9*, e59360. https://doi.org/10.7554/eLife.59360

Bowman, C. R., & Zeithamova, D. (2018). Abstract memory representations in the ventromedial prefrontal cortex and hippocampus support concept generalization. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *38*(10), 2605–2614. https://doi.org/10.1523/JNEUROSCI.2811-17.2018

Bulgarelli, F., & Weiss, D. J. (2016). Anchors aweigh: The impact of overlearning on entrenchment effects in statistical learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(10), 1621–1631. https://doi.org/10.1037/xlm0000263

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. https://doi.org/10.1162/jocn_a_01228

Davachi, L., & DuBrow, S. (2015). How the hippocampus preserves order: The role of prediction and context. *Trends in Cognitive Sciences*, *19*(2), 92–99. https://doi.org/10.1016/j.tics.2014.12.004

Davis, T., Goldwater, M., & Giron, J. (2017). From concrete examples to abstract relations: The rostrolateral prefrontal cortex integrates novel examples into relational categories. *Cerebral Cortex*, *27*(4), 2652–2670. https://doi.org/10.1093/cercor/bhw099

Davis, T., Love, B. C., & Preston, A. R. (2012a). Learning the exception to the rule: Model-based FMRI reveals specialized representations for surprising category members. *Cerebral Cortex*, *22*(2), 260–273. https://doi.org/10.1093/cercor/bhr036

Davis, T., Love, B. C., & Preston, A. R. (2012b). Striatal and hippocampal entropy and recognition signals in category learning: Simultaneous processes revealed by model-based fMRI. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(4), 821–839. https://doi.org/10.1037/a0027865

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

Eichenbaum, H. (2014). Time cells in the hippocampus: A new dimension for mapping memories. *Nature Reviews Neuroscience*, *15*(11), 732–744. https://doi.org/10.1038/nrn3827

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

Endress, A. D., Slone, L. K., & Johnson, S. P. (2020). Statistical learning and memory. *Cognition*, *204*, Article 104346. https://doi.org/10.1016/j.cognition.2020.104346

Ezzyat, Y., & Davachi, L. (2011). What constitutes an episode in episodic memory? *Psychological Science*, *22*(2), 243–252. https://doi.org/10.1177/0956797610393742

Finn, A. S., & Hudson Kam, C. L. (2008). The curse of knowledge: First language knowledge impairs adult learners' use of novel statistics for word segmentation. *Cognition*, *108*(2), 477–499. https://doi.org/10.1016/j.cognition.2008.04.002

Finn, A. S., Kharitonova, M., Holtby, N., & Sheridan, M. A. (2019). Prefrontal and hippocampal structure predict statistical learning ability in early childhood. *Journal of Cognitive Neuroscience*, *31*(1), 126–137. https://doi.org/10.1162/jocn_a_01342

Fiser, J., & Aslin, R. N. (2002). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences*, *99*(24), 15822–15826. https://doi.org/10.4324/9780203770351

Fox, J., & Weisberg, S. (2019). An R companion to applied regression (3rd ed.). Sage. https://socialsciences.mcmaster.ca/jfox/Books/Companion/

Gebhart, A. L., Aslin, R. N., & Newport, E. L. (2009). Changing structures in midstream: Learning along the statistical garden path. *Cognitive Science*, *33*(6), 1087–1116.

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

Giroux, I., & Rey, A. (2009). Lexical and sublexical units in speech perception. *Cognitive Science*, *33*(2), 260–272. https://doi.org/10.1111/j.1551-6709.2009.01012.x

Gomez, R. L. (2016). Do infants retain the statistics of a statistical learning experience? Insights from a developmental cognitive neuroscience perspective. *Philosophical Transactions of the Royal Society B: Biological*

Sciences, *372*(1711), Article 20160054. https://doi.org/10.1098/rstb.2016.0054

Hintzman, D. L. (1984). MINERVA 2: A simulation model of human memory. *Behavior Research Methods, Instruments, & Computers*, *16*(2), 96–101. https://doi.org/10.3758/BF03202365

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

Isbilen, E., McCauley, S. M., Kidd, E., & Christiansen, M. (2017). Testing statistical learning implicitly: A novel chunk-based measure of statistical learning. *The 39th Annual Conference of the Cognitive Science Society* (pp. 564–569). https://doi.org/10.1097/00004032-198811000-00008

Kalm, K., & Norris, D. (2017). A shared representation of order between encoding and recognition in visual short-term memory. *NeuroImage*, *155*, 138–146. https://doi.org/10.1016/j.neuroimage.2017.04.047

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), 12733. 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*, e38550. 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), B35–B42. https://doi.org/10.1016/S0010-0277(02)00004-5

Konkel, A., Warren, D. E., Duff, M. C., Tranel, D. N., & Cohen, N. J. (2008). Hippocampal amnesia impairs all manner of relational memory. *Frontiers in Human Neuroscience*, *2*, 15. https://doi.org/10.3389/neuro.09.015.2008

Kruschke, J. K. (1992). Investigations of an exemplar-based connectionist model of category learning. *Psychology of Learning and Motivation—Advances in Research and Theory*, *28*, 207–250. https://doi.org/10.1016/S0079-7421(08)60491-0

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

Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: A network model of category learning. *Psychological Review*, *111*(2), 309–332. https://doi.org/10.1037/0033-295X.111.2.309

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

Mack, M. L., Love, B. C., & Preston, A. R. (2016). Dynamic updating of hippocampal object representations reflects new conceptual knowledge. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(46), 13203–13208. https://doi.org/10.1073/pnas.1614048113

Mack, M. L., Preston, A. R., & Love, B. C. (2013). Decoding the brain's algorithm for categorization from its neural implementation. *Current Biology*, *23*(20), 2023–2027. https://doi.org/10.1016/j.cub.2013.08.035

Mack, M. L., Preston, A. R., & Love, B. C. (2020). Ventromedial prefrontal cortex compression during concept learning. *Nature Communications*, *11*(1), 46. https://doi.org/10.1038/s41467-019-13930-8

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, *102*(3), 419–457. https://doi.org/10.1037/0033-295X.102.3.419

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

Nosofsky, R. M. (1987). Attention and learning processes in the identification and categorization of integral stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*(1), 87–108. https://doi.org/10.1037/0278-7393.13.1.87

O'Bryan, S. R., Worthy, D. A., Livesey, E. J., & Davis, T. (2018). Model-based fMRI reveals dissimilarity processes underlying base rate neglect. *eLife*, *7*, e36395. https://doi.org/10.7554/eLife.36395

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 & 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), e0125648. https://doi.org/10.1371/journal.pone.0125648

Perruchet, P. (2019). What Mechanisms Underlie Implicit Statistical Learning? Transitional Probabilities Versus Chunks in Language Learning. *Topics in Cognitive Science*, *11*(3), 520–535. https://doi.org/10.1111/tops.12403

Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: One phenomenon, two approaches. *Trends in Cognitive Sciences*, *10*(5), 233–238. https://doi.org/10.1016/j.tics.2006.03.006

Perruchet, P., & Vinter, A. (1998). PARSER: A Model for Word Segmentation. *Journal of Memory and Language*, *39*(2), 246–263. https://doi.org/10.1006/jmla.1998.2576

R Core Team. (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing. https://www.R-project.org/

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

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, *621*(35), 606–621. https://doi.org/10.1006/jmla.1996.0032

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., 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 B: Biological Sciences*. Advance online publication. https://doi.org/10.1098/rstb.2016.0049

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., Guarino, K. F., Schapiro, A. C., Turk-Browne, N. B., & Preston, A. R. (2017). Hippocampal structure predicts statistical learning and associative inference abilities during development. *Journal of Cognitive Neuroscience*, *29*(1), 37–51. https://doi.org/10.1162/jocn_a_01028

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*, 8151. https://doi.org/10.1038/ncomms9151

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM-retrieving effectively from memory. *Psychonomic Bulletin & Review*, *4*(2), 145–166. https://doi.org/10.3758/BF03209391

Siegelman, N., Bogaerts, L., Armstrong, B. C., & Frost, R. (2019). What exactly is learned in visual statistical learning? Insights from Bayesian modeling. *Cognition*, *192*, 104002. https://doi.org/10.1016/j.cognition.2019.06.014

Slone, L. K., & Johnson, S. P. (2018, May). When learning goes beyond statistics: Infants represent visual sequences in terms of chunks. *Cognition*, *178*, 92–102. https://doi.org/10.1016/j.cognition.2018.05.016

Thiessen, E. D. (2017). What's statistical about learning? Insights from modelling statistical learning as a set of memory processes. *Philosophical Transactions of the Royal Society B: Biological Sciences*. Advance online publication. https://doi.org/10.1098/rstb.2016.0056

Thiessen, E. D., & Pavlik, P. I., Jr. (2013). iMinerva: A mathematical model of distributional statistical learning. *Cognitive Science*, *37*(2), 310–343. https://doi.org/10.1111/cogs.12011

Tompary, A., Zhou, W., & Davachi, L. (2020). Schematic memories develop quickly, but are not expressed unless necessary. *Scientific Reports*, *10*(1), 16968. https://doi.org/10.1038/s41598-020-73952-x

Toro, J. M., Sinnett, S., & Soto-Faraco, S. (2005). Speech segmentation by statistical learning depends on attention. *Cognition*, *97*(2), B25–B34. https://doi.org/10.1016/j.cognition.2005.01.006

Treves, A., & Rolls, E. T. (1992). Computational constraints suggest the need for two distinct input systems to the hippocampal CA3 network. *Hippocampus*, *2*(2), 189–199. https://doi.org/10.1002/hipo.450020209

Tummeltshammer, K., Amso, D., French, R. M., & Kirkham, N. Z. (2017). Across space and time: Infants learn from backward and forward visual statistics. *Developmental Science*, *20*(5), e12474. https://doi.org/10.1111/desc.12474

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. https://doi.org/10.1037/0096-3445.134.4.552

Wilburn, C., & Feeney, A. (2008). Do development and learning really decrease memory? On similarity and category-based induction in adults and children. *Cognition*, *106*(3), 1451–1464. https://doi.org/10.1016/j.cognition.2007.04.018

Zacks, J. M., & Swallow, K. M. (2007). Event segmentation. *Current Directions in Psychological Science*, *16*(2), 80–84. https://doi.org/10.1111/j.1467-8721.2007.00480.x

Zaki, S. R., Nosofsky, R. M., Stanton, R. D., & Cohen, A. L. (2003). Prototype and exemplar accounts of category learning and attentional allocation: A reassessment. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(6), 1160–1173. https://doi.org/10.1037/0278-7393.29.6.1160

Zhao, J., & Luo, Y. (2017). Statistical regularities guide the spatial scale of attention. *Attention, Perception & Psychophysics*, *79*(1), 24–30. https://doi.org/10.3758/s13414-016-1233-1

Received July 23, 2020
Revision received May 13, 2021
Accepted May 16, 2021 ■