https://doi.org/10.1037/xge0001091

# Effects of Statistical Learning in Passive and Active Contexts on Reproduction and Recognition of Auditory Sequences

Saloni Krishnan<sup>1</sup>, Daniel Carey<sup>2</sup>, Frederic Dick<sup>3, 4</sup>, and Marcus T. Pearce<sup>5, 6</sup>

<sup>1</sup> Department of Psychology, Royal Holloway, University of London

<sup>2</sup> Novartis Pharmaceuticals, Dublin, Ireland

<sup>3</sup> Department of Psychological Sciences, Birkbeck, University of London

<sup>4</sup> Department of Experimental Psychology, University College London

<sup>5</sup> School of Electronic Engineering and Computer Science, Queen Mary University of London

<sup>6</sup> Department of Clinical Medicine, Aarhus University

Statistical learning plays an important role in acquiring the structure of cultural communication signals such as speech and music, which are both perceived and reproduced. However, statistical learning is typically investigated through passive exposure to structured signals, followed by offline explicit recognition tasks assessing the degree of learning. Such experimental approaches fail to capture statistical learning as it takes place and require post hoc conscious reflection on what is thought to be an implicit process of knowledge acquisition. To better understand the process of statistical learning in active contexts while addressing these shortcomings, we introduce a novel, processing-based measure of statistical learning based on the position of errors in sequence reproduction. Across five experiments, we employed this new technique to assess statistical learning using artificial pure-tone or environmentalsound languages with controlled statistical properties in passive exposure, active reproduction, and explicit recognition tasks. The new error position measure provided a robust, online indicator of statistical learning during reproduction, with little carryover from prior statistical learning via passive exposure and no correlation with recognition-based estimates of statistical learning. Error position effects extended consistently across auditory domains, including sequences of pure tones and environmental sounds. Whereas recall performance showed significant variability across experiments, and little evidence of being improved by statistical learning, the error position effect was highly consistent for all participant groups, including musicians and nonmusicians. We discuss the implications of these results for understanding psychological mechanisms underlying statistical learning and compare the evidence provided by different experimental measures.

Keywords: statistical learning, active sequence reproduction, passive exposure, memory recall, auditory perception

In real-world auditory environments, people are faced with an ever-changing sequence of sounds. The identity of the source, and significance to the listener, are often encoded in statistical regularities that unfold over time. Statistical regularities underlie the acoustics of many natural sound classes (McDermott et al., 2013; Turner, 2010) but are particularly prominent in symbolic cultural

Saloni Krishnan () https://orcid.org/0000-0002-6466-141X Frederic Dick () https://orcid.org/0000-0002-2933-3912

Marcus T. Pearce D https://orcid.org/0000-0002-1282-431X

This research was funded in part by the British Academy and Marie Skłodowska-Curie Actions of the European Commission. The data collected are available at https://osf.io/t7k34/?view\_only=b4df53733f264f91948951f2 63e8eb3d (Krishnan et al., 2020) and were made available as a preprint on PsyRxiv. We thank Adam Tierney for useful comments on a draft of this article and the following for their assistance with data collection: Alexander Sulis, Lina Bergstrom, Amanda Ferrara, Kelli Linder, Nicola Bumpus, and

communication signals, such as bird song, speech, and music. For these cultural communicative signals, statistical learning plays an important role in the acquisition of internal cognitive representations of signal structure (Christiansen & Chater, 2016; Hay et al., 2011; Huron, 2006; Pearce, 2018; Romberg & Saffran, 2010; Takahasi et al., 2010; Wells et al., 2009). Cultural communicative

Elizabeth Brown. We would also like to acknowledge the work of Marcia Chew, whose master's thesis provided the data for Experiment 4.

Correspondence concerning this article should be addressed to Saloni Krishnan, Department of Psychology, Royal Holloway, University of London, Egham Hill, Egham, Surrey, TW20 0EX, United Kingdom, or Frederic Dick, Department of Psychological Sciences, Birkbeck, University of London, Malet Street, London, WC1E 7HX, United Kingdom, or Marcus T. Pearce, School of Electronic Engineering and Computer Science, Queen Mary University of London, Mile End Road, London E1 4NS, United Kingdom. Email: Saloni.Krishnan@rhul.ac.uk

signals such as speech and music are further distinguished from other natural sensory phenomena by the fact that we not only perceive but also reproduce them, especially during learning. Accordingly, research increasingly focuses on integrated theories encompassing both comprehension and production (e.g., Pickering & Garrod, 2013). Yet, as we outline below, most evaluations of human statistical learning have used post hoc measures of perception to evaluate learning of signal structure. In the present research, we introduce and evaluate a new measure of *dynamic statistical learning* based on errors made while reproducing sound sequences with embedded statistical regularities. We use these error patterns to evaluate combined effects of passive exposure and active reproduction of statistically structured information and compare this new approach to existing measures of statistical learning.

Statistical learning is thought to represent a largely implicit process of acquiring an internal cognitive model of the statistical regularities underlying an environmental domain. Contemporary accounts conceive of statistical learning as arising from underlying processes of learning and memory such as chunking (Christiansen, 2019; Perruchet, 2019).<sup>1</sup> More specifically, Christiansen (2019) proposed that statistically based chunking provides a way of overcoming the now-or-never bottleneck (Christiansen & Chater, 2016), arising from a limited ability to process and store information that arrives sequentially, with each element being perceived only for a fleeting instant. According to this account, learning of statistical regularities in the input can allow commonly co-occurring elements to be chunked and passed on as more abstract units for use in higher-level chunk-and-pass processing. There is increasing evidence of a close relationship between statistical learning and chunking, although it remains unclear whether one precedes the other (Perruchet, 2019; Perruchet & Pacton, 2006) or the two processes operate in parallel (Isbilen et al., 2020; McCauley & Christiansen, 2019).

Much of the literature on *statistical learning* relies on measures referred to as reflection-based by Christiansen (2019), most typically a comparative familiarity judgment following passive exposure to a structured auditory or visual stream. For instance, Saffran, Newport, and Aslin (1996) passively exposed adults to sequences of syllables constructed by concatenating triplets of syllables (dubbed words) such that boundaries between words were indicated by lower transitional probabilities than syllable transitions occurring within words. Statistical learning of the probabilistic structure of these sequences was subsequently assessed using an explicit two-alternative forced-choice (2AFC) recognition task discriminating words from either nonwords or part-words differing by one syllable. The results showed above-chance performance for recognizing words against both nonwords and part-words, with an additional analysis showing better performance for words exhibiting greater compared to lower transitional probabilities.

This basic statistical learning paradigm—passive exposure followed by a 2AFC posttest—has been highly influential and demonstrated sensitivity to statistical properties of pitch sequences (Saffran et al., 1999), pitch interval sequences (Saffran & Griepentrog, 2001; Saffran et al., 2005), and synthesized instrumental timbres (Tillmann & McAdams, 2004). Under certain conditions, nonadjacent statistical dependencies can be learned in sequences of syllables (Newport & Aslin, 2004) and tones (Creel et al., 2004). Statistical learning has also been demonstrated for sequences of abstract visual shapes (Fiser & Aslin, 2002; Kirkham et al., 2002), audio-visual sequences (Mitchel & Weiss, 2011), and sequential patterns of tactile finger stimulation (Conway & Christiansen, 2005). In spite of this apparent domain generality, differences in statistical learning have been observed across domains (Conway & Christiansen, 2006; Emberson et al., 2011; Johansson, 2009; Saffran, 2002). This has led to a proposal that statistical learning reflects a set of domain-general mechanisms with modality-specific neural implementations subject to different biases (Frost et al., 2015), although more recent work (Siegelman, Bogaerts, Elazar, et al., 2018) suggests that at least some of these modality differences may instead reflect previous experience with the class of stimuli, particularly speech.

According to a recent estimate, the 2AFC recognition task was used in 60% of all studies on statistical learning between 1996 and 2016 (Frost et al., 2019). However, this task suffers from distinct and severe disadvantages as a measure of statistical learning (Christiansen, 2019; Frost et al., 2019; Isbilen et al., 2020; Siegelman et al., 2017). First, as an offline measure recorded after learning has taken place, it provides no insight into the dynamic time course of learning (i.e., rate and trajectory) nor integration with knowledge acquired through prior learning. Second, the binary nature of the 2AFC task provides a limited window into the fine-grained nature of the knowledge acquired, and while performance is typically above chance across a sample, a large minority of participants typically perform at chance level. Finally, rather than directly accessing the psychological mechanisms involved in implicit statistical learning, 2AFC recognition performance is an indirect measure that relies on conscious reflection about what has been learned, which may add significant noise and underestimate effects of statistical learning. Importantly, these limitations are not only methodological, but they also place significant limitations on our theoretical understanding of the psychological representations and processes involved in statistical learning (Siegelman, Bogaerts, et al., 2018).

As a result of these limitations, Christiansen (2019) argued the case for greater use of what he calls processing-based measures, often used in research on implicit learning (Reber, 1967, 1989), which allow online investigation of statistical learning dynamically as it is taking place. Perhaps the most prominent example of such a measure is the serial reaction time (RT) task (SRT; Nissen & Bullemer, 1987) often used as a measure of motor sequence learning (e.g., Willingham et al., 2000). In the SRT, participants respond to each element in a sequence by pressing the corresponding button in a spatialized array as quickly as possible. The key relevant SRT finding is that participants' RTs become faster for deterministically repeating sequences than for random sequences (Robertson, 2007). Hunt and Aslin (2001, 2010) extended the SRT to statistically structured visual sequences along the same lines as those used by Saffran, Newport, and Aslin (1996), finding faster reaction times to the second and third elements of a triplet than to the first (which had a lower transitional probability) and showing that these effects depend on computation of both joint bigram probability and conditional bigram probability.

In the auditory domain, Misyak et al. (2010a) used an SRT task to examine learning nonadjacent dependencies in sequences of artificial words, finding a trajectory of decreasing RTs for predictable compared to unpredictable words throughout training,

<sup>&</sup>lt;sup>1</sup> Here, "chunk" is defined as a cognitive grouping or unit composed of subelements such as tones or syllables (Perruchet & Pacton, 2006; Perruchet, 2019).

followed by an increase when presented with a block of ungrammatical sequences, then recovery of the trajectory for a final block of grammatical items. The learning trajectory was successfully simulated by a cohort of 30 simple recurrent networks (SRNs; Elman, 1990), each with different random initialization (Misyak et al., 2010b). Franco and Destrebecgz (2012) noted that SRT learning could reflect an associative process of conditional probability estimation (represented computationally by the SRN) or, alternatively, chunking in memory (represented by models such as PARSER; Perruchet & Vinter, 1998). They attempted to tease apart these accounts by presenting two blocks of a visual SRT task with distinct grammars L1 and L2, which were either disjoint or conflicting, followed by a recognition task presenting L2 words, L2 part-words, and nonwords. Consistent with associative learning rather than chunking, recognition performance did not differ between the disjoint and conflicting conditions, and performance for part-words was better than for nonwords.

Siegelman, Bogaerts, et al. (2018) developed a processingbased measure related to the SRT in which participants self-pace through a statistically structured set of visual sequences before undertaking a traditional recognition task. Statistical learning was indicated by a trajectory of faster RTs for predictable (with high transitional probability) than unpredictable elements. The rate of learning was faster for elements with greater transitional probability and slower when preceded by learning a different grammar for the same visual shapes, highlighting the advantages of processingbased measures. The primary limitation of the SRT task and related processing-based measures of statistical learning based on RTs is their low signal-to-noise ratio.

Another processing-based measure of statistical learning is the Hebb repetition effect (HRE; Hebb, 1961), a serial recall task in which one of the set of sequences to be recalled is covertly repeated periodically throughout a session, leading to better recall performance for the repeated sequence than the filler sequences (Page et al., 2006; Page & Norris, 2009). This neatly isolates the ability to learn repeated sequences from general short-term memory (STM) capacity and provides an online measure of learning. An interesting phenomenon that has emerged from research on the HRE is that errors made during recall are themselves learned, in some cases cancelling out learning of the repeated sequence itself (Lafond et al., 2010). This highlights that fact that learning takes place during both presentation and recall of the Hebb sequence. Rey et al. (2020) introduced a variant of the HRE in which participants read aloud a sequence of visually presented letters, in which a triplet of letters reoccurred periodically. The results showed a progressive reduction in speech onset latency for the predictable second and third letters in the Hebb sequence but only when the filler sequences were sufficiently differentiated (i.e., by using consonants compared to vowels), noisy (i.e., with increased alphabet size), or short with respect to the length of the Hebb sequence. This suggests that learning from the filler sequences can disrupt detection of the Hebb sequence. Furthermore, there was evidence of faster learning for the second than the third letter of the sequence (see also Kim et al., 2009; Turk-Browne et al., 2005), suggesting an associative mechanism of conditional probability estimation, represented computationally by the SRN, rather than a chunking mechanism, represented by models such as PARSER and TRACX (French et al., 2011; Mareschal & French, 2017). One disadvantage of the HRE is that it has low test-retest reliability, resulting partly from the fact that it relies on comparing recall performance between the filler and Hebb sequences (Bogaerts et al., 2018).

In the present work, we introduce and evaluate a new processing-based measure of statistical learning based on errors in memory recall. This contrasts with research using the SRT (Misyak et al., 2010a) and target detection tasks (Batterink et al., 2015; Franco et al., 2015), which do not require participants to memorize and recall sequences, and with the Hebb repetition paradigm, which does not present sequences containing internal statistical regularities but rather creates regularity by repeating entire sequences throughout the experimental session. We focus on recall of statistically structured sequences for two reasons, both alluded to above. First, learning of many cultural systems such as language and music is based not on mere exposure but on active reproduction through imitation (Heyes, 2018). This motivates an examination of learning during active reproduction in its own right but also in comparison to the learning that takes place during passive exposure, as investigated by the bulk of research in statistical learning. Second, as noted above, it is thought that an intimate relationship exists between implicit statistical learning and basic processes in memory encoding and storage such as chunking (Christiansen, 2019; Perruchet, 2019). Therefore, a memory-based task such as recalling a statistically structured sequence may yield a very direct, processing-based probe of statistical learning as it is actually happening.

Research using sequence reproduction to investigate implicit learning goes back to Miller (1958). In his experiment, participants learned letter strings through repeated exposure and recall. The letter strings were generated either without an overt structure or via an artificial grammar. Miller found that recall performance was better for the grammatical strings (see also Redington & Chater, 2002). Reber (1967) also reported better recall performance for grammatical compared to random items. Somewhat more recently, Karpicke and Pisoni (2004) presented participants with a reproduction task for sequences generated by one of two artificial grammars, differing only in the mapping from spatial position to colors. They compared three conditions: visual only (colored lights), verbal labels only (spoken color names), and multimodal (lights and names). Across all three conditions, reproduction in a subsequent test phase was better for sequences from the trained versus untrained grammar. In a subsequent recognition memory test, participants were more likely to judge sequences from the trained grammar as familiar than sequences from the untrained grammar, regardless of whether the sequence was actually old or new. While performance in the recognition memory task was better for items that contained frequently occurring bigrams and trigrams, there was no difference in recall performance for these items. In a closely related study, Conway et al. (2010; Experiment 2) also found better reproduction performance for grammatical than nongrammatical stimuli on a reproduction task. Here, Conway et al. used verbal recall of auditory sequences composed of four artificial syllables generated by an artificial grammar.

Notably, these auditory implicit learning studies have not compared effects of active reproduction with passive exposure, nor have they examined dynamic assessment of progressively increasing acquisition of knowledge during recall-based statistical learning tasks. Doing so in the present research allows us to investigate whether knowledge acquired during passive exposure shares overlapping representations with knowledge acquired during recall (see Lafond et al., 2010) and better understand the trajectory of learning through recall of auditory sequences (see Siegelman, Bogaerts, et al., 2018).

If processing-based measures do indeed reflect more directly and accurately the underlying psychological mechanisms of statistical learning, we would expect to observe them diverging significantly from reflection-based measures. In other words, individuals who show strong evidence of statistical learning on processingbased measures should not necessarily do so on reflection-based measures (and vice versa). A handful of studies have addressed this question in the visual (Bertels et al., 2012; Isbilen et al., 2020; Kim et al., 2009; Siegelman, Bogaerts, Kronenfeld, & Frost, 2018) and auditory (Batterink et al., 2015; Franco et al., 2015; Isbilen et al., 2020; Misyak et al., 2010a) domains. The results are somewhat mixed, with some studies finding significant correlations (Isbilen et al., 2020; Siegelman, Bogaerts, Kronenfeld, & Frost, 2018) but the majority finding no relationship. The question therefore remains unresolved, especially for auditory recall paradigms (Isbilen et al., 2020), where it has not been addressed at all for nonlinguistic stimuli.

The present research examined statistical learning during sequence reproduction using auditory sequences with systematically controlled statistical structure in three ordered contexts: passive exposure, active sequence reproduction, and explicit sequence recognition. We investigated in detail the relationships between an existing processing-based measure (sequence recall), a new dynamic processing-based measure based on error position effects during reproduction, and the most widely used reflection-based measure of statistical learning (2AFC sequence recognition). Most importantly, we asked how these three measures might differentially index statistical learning from both passive exposure and active reproduction. Combining these approaches allowed us to investigate the effects of implicit statistical learning on auditory sequence reproduction, affording a finer-grained analysis of the representations that are learned at the level of individual events. We expected our new processing-based measure of error position effects to show more detailed sensitivity to the statistical properties of the stimuli and to allow for dynamic tracking of statistical learning over time (cf. Karpicke & Pisoni, 2004).

For the active sequence reproduction task, like Karpicke and Pisoni (2004) and Conway et al. (2010), we adapted the electronic children's toy "Simon" from the 1970s. Simon is a simple game device that produces sequences of lighted buttons with corresponding tones. The user has to reproduce each sequence using buttons on the device. The length of the sequence is increased by one when the entire sequence is successfully reproduced and terminated whenever an error is made. This game has been used to probe STM in various populations including cochlear implant patients (Cleary et al., 2001), 8–9-year-old children (Pisoni & Cleary, 2004), older and younger adults with impaired hearing (Humes & Floyd, 2005), and musicians (Tierney et al., 2008). Generally, performance on the Simon task is better in the audiovisual condition than either unimodal condition.

In the set of five experiments reported here, all participants completed an active sequence reproduction task (Simon) using sequences from an artificial language (Experiment 1a being the only exception). In most experiments, before undertaking the Simon task, participants were first familiarized with one of the artificial languages through passive exposure to artificially constructed pure-tone sequences with systematically controlled statistical structure (Experiments 1b, 4, and 5 being the exceptions). After the Simon task, participants completed a 2AFC sequence recognition task (all experiments except Experiment 5).

In Experiment 1, we examined whether active reproduction of statistically structured sequences results in statistical learning, as assessed by our new processing-based measure based on analysis of error positions (Experiment 1b). We also confirmed, for the same stimulus materials, that learning takes place using the traditional 2AFC recognition task following both passive exposure (Experiment 1a) and active reproduction (Experiment 1b). Having demonstrated that the error position effect captures statistical learning, we examined how the measure can be deployed to assess statistical learning in dynamic situations. Specifically, Experiment 2 examined the effects on statistical learning of combining passive exposure and active reproduction, assessed using both the reflection-based 2AFC recognition task and processing-based error position effects in reproduction. In doing so, we assessed whether these measures show improved learning when reproduction is preceded by passive exposure and whether this depends on the length of the exposure. We also probed whether acquired knowledge is carried over from passive exposure to active reproduction.

Experiments 3, 4, and 5 were control experiments. Experiment 3 controlled for the possibility that the effects observed in Experiments 1 and 2 reflect some artifact of the experimental design by using randomly generated unstructured sequences in either (or both) the passive familiarization and active reproduction tasks. Experiment 4 controlled for the possibility that the effects observed in Experiments 1 and 2 were driven by the colored lights that accompanied each tone sequence in the Simon reproduction task, while also investigating whether the error position effect extends to sequences of environmental sounds and to participants with musical training. Experiment 5 controlled for the possibility that the effects observed in Experiments 1 and 2 were driven by the spatial assignment of tones to buttons in the reproduction task by randomizing the tone-to-button assignment across participants. Specific hypotheses are presented in the introductions to each experiment.

The remainder of the article is organized as follows: the next section introduces the general method common to all experiments, followed by sections presenting additional methods, results, and cumulative discussion for Experiments 1 to 5. The final section presents a general discussion of the results and their implications for future research. The experimental data are available at https://osf.io/t7k34/?view\_only=b4df53733f264f91948951f263e8eb3d.

# **General Method**

# **Participants**

For Experiments 1–4, participants were recruited using the Birkbeck and University College London participant pools. They were right-handed, were aged from 16 to 45, and reported no history of speech, language, or hearing difficulties. Other than the musicians in Experiment 4, all participants reported having no prior musical training, and none reported having absolute pitch. In Experiment 5, participants were recruited online (see below). All

experiments received prior ethical approval from the Birkbeck Research Ethics Committee.

The minimum number of participants per condition was determined by estimates of effect size in the literature. For recognition tests following passive exposure, Saffran et al. (1999) reported a large effect size ( $d \sim 1.2$ ); when assuming a slightly smaller effect size (d = .8), a sample size of 15 gives 90% power to detect an effect of statistical structure on recognition when using a onetailed *t*-test to test difference from a constant. For detecting effects of statistical learning using reproduced sequence length, Tierney et al. (2008) reported a large effect of musical expertise on auditory sequence learning using the Simon task (d = 2.11). Given the current paradigm is much shorter, we thus used an effect size of half that (d = 1.05) for power analysis, which indicated a sample size of 17 in each group would provide 90% power to detect differences in sequence lengths (using one-tailed t-tests to test differences between two independent means). Therefore, we collected samples of 17+ participants in all experiments. In most cases (all but Experiment 3a), we overrecruited due to higher participant uptake than anticipated. Recruitment was always terminated at 30 participants. For the online study (Experiment 5), there were 24 possible tone-to-button assignments, so we recruited 48 participants (24 assigned to each language).

# Stimuli

### **Passive Familiarization**

Following previous research (e.g., Saffran et al., 1999), a *structured artificial pure-tone language* was created, consisting of isochronous tone sequences using an alphabet of four tones. These tones were sinusoids of 300-ms duration that formed the notes of a C-major chord (262 Hz [C4]; 327.5 Hz [E4]; 393 Hz [G4]; 524 Hz [C5]). All tones were normalized for equal root-mean-square amplitude and had 50-ms onset and offset ramps; there was no additional gap between tones.

The artificial language consisted of four tone words (sequences of three tones), which were chosen so that when randomly concatenated together, the longer sequences thus created would have particular statistical properties. In particular, each of the four tones would be equiprobable overall (i.e., a uniform zerothorder probability of occurrence), but the first-order transition (or bigram) probabilities between tones would be greater within than between tone words. For the familiarization phase of passive exposure, an artificial language block of 690 tones was created by randomly concatenating 230 tone words, subject to two conditions: First, no tone word appeared twice consecutively; second, within-word bigrams never appeared between tone words. As a result of these constraints, each tone word could be followed by one of two other tone words. A second structured artificial language (L2) was created by reversing the tone words of the first (L1), and the same process applied to create familiarization sequences for passive exposure. We use the terms fami*liarized language* and *nonfamiliarized language* to refer to the artificial language presented and not presented, respectively, during passive familiarization.

A third artificial language was also constructed, for use as a control, by creating sequences in which the four tones appear with equal probability. We refer to this as the *unstructured language*. The unstructured language has the same uniform zeroth-order probabilities as L1 and L2 but is also uniform at the level of firstorder (and higher) transitional probabilities (i.e., every tone has an equal probability of occurrence regardless of the preceding tone). For the familiarization phase, a block consisted of 690 tones from the unstructured language.

Table 1 shows the tone words, within-word transitions, and between-word transitions for L1, L2, and the unstructured language. Table 2 gives the corresponding transition matrices.

# Active Reproduction

Stimuli for the Simon active reproduction task were 10 distinct sequences created by concatenating seven consecutive tone words from L1, L2, or the unstructured language. Each sequence thus had 21 tones.

### Sequence Recognition

A 2AFC test was used to assess statistical learning of the probabilistic structure of the tone sequences previously presented. Tone words from L1 were paired exhaustively with tone words from L2 to create 16 distinct pairs of tone words. Each pair was presented three times, producing 48 trials in total, except for Experiment 4, as described in the experiment-specific procedure section.

Two randomized orders of presentation were constructed. The order of presentation of the two tone-words in a pair was counterbalanced between orders such that a pair presented with L1 first and L2 second in one order appeared with L2 first and L1 second in the other order. The use of the two orders of presentation was counterbalanced across participants and artificial language used for passive familiarization.

### **Apparatus and Procedure**

Each experiment could optionally include three components: (a) a phase of passive familiarization through exposure to tone sequences, (b) an active sequence reproduction task, and (c) a 2AFC sequence recognition task. Not every experiment included all three components, but in all experiments, components were presented in the order given above. The presence and nature of each component for each experiment is summarized in Figure 1.

Except for Experiments 4 and 5, data were collected using a second-generation iPad with custom-designed apps using the touchpad for the reproduction and recognition tasks, and Sennheiser HD-25 headphones were used to present the auditory stimuli. Experiment 4 instead used a laptop and gamepad, while Experiment 5 was run online, as described in the experiment-specific procedure sections.

### **Passive Familiarization**

Participants were exposed to a continuous sequence of tones via headphones. One block consisted of 690 tones, lasting about 3.5 min. Depending on the experiment (see Figure 1), L1, L2, or the unstructured language was presented in one or three blocks. Participants were informed they would be listening to a stream of sounds but that they did not need to pay attention to them and were not informed about the presence of any regularities in the stimulus. During this time, participants

Table 1		
The Structured and	Unstructured Ar	tificial Languages

Description	L1	L1 TP	L2	L2 TP	Unstructured language TP
Tone words	C4-E4-G4		G4-E4-C4		
	E4-G4-C5		C5-G4-E4		
	G4-C5-C4		C4-C5-G4		
	C5-C4-E4		E4-C4-C5		
Within-word transitions (high probability)	$C4 \rightarrow E4$	.667	C4→C5	.667	.25
	$E4 \rightarrow G4$	.667	E4→C4	.667	.25
	$G4\rightarrow C5$	.667	$G4 \rightarrow E4$	.667	.25
	C5→C4	.667	C5→G4	.667	.25
Between-word transitions (low probability)	C4→C5	.167	C4→E4	.167	.25
	C4→C4	.167	C4→C4	.167	.25
	E4→C4	.167	$E4 \rightarrow G4$	.167	.25
	$E4 \rightarrow E4$	.167	$E4 \rightarrow E4$	.167	.25
	$G4 \rightarrow E4$	.167	G4→C5	.167	.25
	$G4 \rightarrow G4$	.167	$G4 \rightarrow G4$	.167	.25
	C5→G4	.167	C5→C4	.167	.25
	C5→C5	.167	C5→C5	.167	.25
Other transitions	C4→G4	0	C4→G4	0	.25
	$E4 \rightarrow C5$	0	E4→C5	0	.25
	G4→C4	0	G4→C4	0	.25
	$C5 \rightarrow E4$	0	$C5 \rightarrow E4$	0	.25

*Note.* The table shows the tone words of Structured Artificial Languages 1 (L1) and 2 (L2), the within-word tone transitions with high transitional probability (TP), the between-word tone transitions with low TP, and finally tone transitions that never appeared in the stimulus sequences. The final column shows the corresponding (uniform) first-order probabilities in the unstructured language. Recurring decimals have been rounded to three decimal places.

in Experiments 1–3 completed the Edinburgh Handedness Inventory and a short questionnaire about their natural language background.

### Active Reproduction

Participants performed an audio-visual sequence reproduction task, similar to the Simon game (Cleary et al., 2001; Karpicke & Pisoni, 2004). For each sequence presented, participants experienced a number of trials, with the sequence length increasing by one on each trial until either a reproduction error was made or the end of the sequence was reached. The gaming interface comprised an arrangement of four buttons each with a different color, where colored buttons were arranged in a circular configuration (see Figure 2). Four tones (262 Hz [C4]; 327.5 Hz [E4]; 393 Hz [G4]; 524 Hz [C5]) were each uniquely paired with a single button (see Figure 2 for pairings). Tone/button pairings were randomly set prior to data collection and were fixed throughout Experiments 1–4. Except in Experiment 4, each button was illuminated, and the associated tone was played simultaneously during exposure and reproduction.

The task on each trial was to listen to the sounds presented by the interface and then, following a cue (an icon presented at the

### Table 2

Transition Matrices for Structured Artificial Language 1 (Top), Structured Artificial Language 2 (Middle), and the Unstructured Language (Bottom)

	C4	E4	G4	C5
<u>C4</u>	.167	.667	0	.167
E4	.167	.167	.667	0
G4	0	.167	.167	.667
C5	.667	0	.167	.167
C4	.167	.167	0	.667
E4	.667	.167	.167	0
G4	0	.667	.167	.167
C5	.167	0	.667	.167
C4	.25	.25	.25	.25
E4	.25	.25	.25	.25
G4	.25	.25	.25	.25
C5	.25	.25	.25	.25

*Note.* Each cell represents the probability of a transition from the tone listed for the corresponding row to the tone listed for the corresponding column. Recurring decimals have been rounded to three decimal places.

A Summary of the Experimental Conditions Differentiating Each of the Experiments Reported Below



*Note.* Where present (all experiments except Experiments 1b, 4, and 5), passive familiarization always appeared first, followed by the active reproduction task where present (all experiments except Experiment 1a), followed by the sequence recognition task where present (all experiments except Experiment 5). L = language.

center of the display), reproduce the sounds in the same order by pressing the corresponding buttons on the interface. The first trial of a sequence always started with one tone. If all the sounds presented in a trial were reproduced correctly, on the next trial, the sequence length was incremented by one tone. As soon as an error was committed in a trial, the trial immediately terminated, and the participant's score for that sequence was recorded as the *sequence length* in the previous trial (which by definition had to have been correctly reproduced in its entirety). After an incorrect response, the next trial presented the first element of a new sequence. The goal of the task was to achieve as high a score as possible. The score for each sequence was displayed on the bottom of the screen and was updated after every trial.

A response time of 3 s was allowed between each button press. If no response was recorded during this time, the experiment automatically moved to the next sequence, recording a missing response for the current sequence. To introduce participants gently to each sequence while not making the experiment too long and fatiguing, for Experiments 1–4, the interstimulus interval (ISI) between items varied according to sequence length. ISI was always consistent within a trial but varied across trials according to sequence length as follows: length 1–3: 500 ms; length 4–6: 400 ms; length 7–12: 200 ms. From length 13–17, the ISI decreased by 10 ms for each tone added until it reached a constant value of 150 ms from length 17–21, producing interonset intervals ranging from 800 to 450 ms.

A sequence was terminated when the participant reproduced the full sequence length of 21 items or when they reproduced the sequence incompletely or incorrectly. Participants could then take a short break before beginning the next sequence. A message was displayed after termination of a sequence. The next trial started when the participant pressed the button labeled "Play next" on this message.

At the beginning of each experiment, participants completed three practice sequences of six items. If a participant reproduced less than four items on either practice sequence, the three practice sequences were rerun until a minimum of four items were successfully reproduced for at least two practice sequences. The 10 test sequences were then presented. Sequences were composed from L1, L2, or the unstructured language depending on the experiment.

#### Sequence Recognition

A 2AFC task was used to assess sequence recognition in Experiments 1–4 with a visual display showing two loudspeakers enabling participants to respond. Both speakers were illuminated when a stimulus was played. Participants were instructed to respond at the end of each trial by pressing one of the speakers, corresponding to the stimulus played first (left) or second (right).

Two tone words were played one after the other, and participants were then asked to select the one they felt was most familiar based on the tasks they had just completed. Prior to undertaking the recognition task, participants were familiarized with the response buttons; the left button was associated with the tone word that came first, and the right button with the tone word that was presented second. Participants responded by pressing the appropriate button; this also

**Figure 2** *The Interface Used for the Simon Reproduction Task* 



*Note.* Panel A: The display used for the custom iPad version of the Simon task (Experiments 1–3). The musical note values (C4, E4, etc.) show the tone frequency paired with that button. Panel B: The display used for the laptop and online versions of the task (Experiments 4–5). For Experiment 4 with the laptop, the gamepad shown in the bottom right was used, and musical notes show the tone frequency used for the tone condition (environmental sound screen not shown). For online Experiment 5, participants used a keyboard to respond (not shown); tone frequency to button mapping was different for each participant (see Experiment 5, "Participants and Procedure"). See the online article for the color version of this figure.

advanced the experiment to the next trial. Participants heard three practice trials prior to testing but were given no feedback about their choices.

The tone words were separated by an ISI of .75 s. The maximum intertrial interval was 3 s but could be shorter depending on the participant's RT to the previous trial.

For all recognition trials, one of the tone words was from L1, and the other was from L2. For participants exposed to L1, the correct choice was the tone word from L1, while for participants exposed to L2, the correct choice was the tone word from L2.

### **Experiment 1**

The purpose of this experiment was to examine whether the positions of errors made in the sequence reproduction task reflected online statistical learning of sequential regularities during the task itself. We hypothesized that, with increasing exposure, errors would increasingly occur disproportionately at points of low transitional probability (i.e., at between-word positions). We also hypothesized that both passive exposure (Experiment 1a) and active reproduction (Experiment 1b) of the same structured tone sequences would induce statistical learning as assessed by subsequent recognition performance.

# **Participants and Procedure**

Forty-one participants (*M* age = 22.02, *SD* = 3.76) completed the experiment, randomly assigned to one of two conditions. In Experiment 1a (familiarization only), participants (*N* = 20, *M* age = 21.35, *SD* = 3.13, 14 female, two undisclosed) were passively exposed to three blocks ( $\sim$ 10.5 min) of structured tone sequences, counterbalanced across participants. They then completed the recognition task (without first performing the reproduction task). In Experiment 1b (reproduction only), participants (N = 21, M age = 22.67, SD = 4.26, 10 female) received no passive familiarization but completed the active reproduction task with 10 structured sequences, counterbalanced across participants, followed by the recognition task for the corresponding artificial language.

### **Results and Discussion**

### Experiment 1a (Familiarization Only)

For the recognition task, tone words from the familiarized language were chosen over those from the nonfamiliarized language on 61.88% of trials (95% CI [56.74, 67.01]), significantly greater than chance (50%), t(19) = 4.54, p < .001, Hedges' g = 1.41 (see Figure 3). This replicates previous findings (e.g., Saffran et al., 1999) for the present stimulus materials.

### Experiment 1b (Reproduction Only)

For the recognition task, tone words from the language used in the reproduction task were chosen on 62.1% of trials (95% CI [58.46, 65.75]), significantly greater than chance performance (50%), t(20) = 6.50, p < .001, g = 1.97. Furthermore, there was no significant difference in recognition performance between Experiments 1a and 1b, t(39) = .07, p = .94, g = .02, suggesting that active sequence reproduction in Experiment 1b promoted statistical learning to a similar extent as passive familiarization in Experiment 1a. Recognition performance for both experiments is plotted in Figure 3.

### Figure 3

Mean Proportions of Correct Responses for the Recognition Task in Experiments 1a (Familiarization Only) and 1b (Reproduction Only)



*Note.* Chance performance is at 0.5, shown by the horizontal line, and asterisks indicate performance differing significantly from chance. Error bars represent 95% confidence intervals around the mean.

In the reproduction task itself, we examined error positions within the sequence by distinguishing sequences with errors between two consecutive tone words versus within a tone word. We defined between-word errors as those occurring on the first element of a tone word such that the (lower) transitional probability spans the boundary between the previous and new word. We defined within-word errors as those committed when reproducing the second or third element of a tone word. We excluded sequences in which an error was made on the first sequence element (where there is no prior context), as well as those where no errors were made (i.e., the participants reproduced an entire sequence of 21 elements). Data were missing for two sequences. In the remaining data, there were 100 sequences containing a between-word error (48%), 62 sequences containing a within-word error (30%), 37 sequences containing an error on the first element (18%),<sup>2</sup> and nine sequences containing no errors (4%). After removing sequences with no errors and sequences with errors on the first element, the pattern of error positions (between:within, 62:38) was significantly different from that predicted by chance (between:within, 30:70),  $\chi^2(1) = 77.66$ , p < .001, V = .61.<sup>3</sup> Participants were significantly more likely to make errors between tone words than within a tone word, suggesting an online effect of dynamic statistical learning of regularities during the active reproduction task (see Figure 4).

Finally, we turn to recall performance. Participants correctly reproduced an average of 10.40 elements (95% CI [9.47, 11.32]).

### Figure 4

Relative Frequency of Errors Committed Between Tone Words in the Reproduction Task for Experiment 1b (Reproduction Only)



*Note.* Chance performance is shown in the rightmost bar, and asterisks indicate performance differing significantly from chance.

# **Experiment 2**

Experiment 1 showed that active sequence reproduction engages statistical learning mechanisms, which can be assessed by error position effects during performance of the task. To investigate whether passive exposure before active reproduction leads to enhanced statistical learning, Experiment 2 combined initial passive exposure of varying durations with active sequence reproduction; statistical learning was again assessed via error position effects and recognition performance.

We hypothesized two effects: (a) processing-based measures of statistical learning during sequence reproduction (recall and error position effects) would be facilitated by prior passive exposure to structured sequences (Experiment 2a), and this facilitation would be less effective with a shorter period of exposure (Experiment 2b) and (b) recognition performance would improve when participants both passively listened to *and* actively reproduced structured sequences, compared with either passive exposure or active reproduction alone.

### **Participants and Procedure**

Sixty participants (*M* age = 25.18, *SD* = 6.40, 29 female) completed the experiment and were randomly assigned to one of two conditions. In Experiment 2a (familiarization + reproduction), participants (*N* = 30, *M* age = 24.23, *SD* = 5.81, 17 female) were passively familiarized with L1 or L2 (counterbalanced), just as in Experiment 1a. They then undertook the reproduction task (as in Experiment 1b) before completing the recognition task. The procedure for Experiment 2b (short familiarization + reproduction; *N* = 30, *M* age = 26.13, *SD* = 6.90, 14 female) was exactly the same, except that passive familiarization consisted only of one block (~3.5 min) rather than three blocks. Participants in Experiment 2b completed the Raven's progressive matrices during the passive familiarization in addition to the handedness inventory and language questionnaire completed during exposure in all experiments (see "General Method").

# **Results and Discussion**

# Experiment 2a (Long Familiarization + Reproduction)

Participants chose tone words from the familiarized language over those from the nonfamiliarized language on 59.03% of test trials (95% CI [54.93, 63.13]), at a level significantly greater than chance, t(29) = 4.32, p < .001, g = 1.1 (see Figure 5). Performance did not differ significantly from that following passive familiarization only (Experiment 1a), t(48) = .85, p = .40, g = .24, demonstrating that including an active reproduction task between passive familiarization and sequence recognition had no additional impact (either beneficial or detrimental) on recognition performance. Furthermore, performance did not differ

<sup>&</sup>lt;sup>2</sup> The relatively large number of errors made at the first sequence element might at first seem surprising, but recall that such errors can be made after listening to a sequence of any length up to 20 tones.

 $<sup>^{3}</sup>$  After removing the first tone, there were 20 positions at which errors could be made, six of which corresponded to between-word positions (30%), while the remaining 14 corresponded to within-word positions (70%).

Mean Proportions of Correct Responses for the Recognition Task in Experiment 2a (Familiarization + Reproduction) and 2b (Short Familiarization + Reproduction)



*Note.* Chance performance is at 0.5, shown by the horizontal line, and asterisks indicate performance differing significantly from chance. Error bars represent 95% confidence intervals around the mean.

significantly from that following active reproduction only (Experiment 1b), t(49) = 1.04, p = .30, g = .29, demonstrating that adding passive familiarization had no additional impact on recognition performance over and above the effect of statistical learning in the reproduction task.

On the reproduction task, data were missing for two sequences, and in the remaining data, there were 122 sequences containing a between-word error (41%), 126 sequences containing a withinword error (42%), 40 sequences containing an error on the first element (13%), and 10 sequences containing no errors (3%). The proportion of errors made between relative to within tone words (49:51) was significantly different from that predicted by chance (30:70),  $\chi^2(1) = 43.51$ , p < .001, V = .38 (see Figure 6).

Finally, we turn to recall performance. Participants correctly reproduced an average of 8.85 elements (95% CI [7.80, 9.89]).

#### *Experiment 2b* (Short Familiarization + Reproduction)

Participants chose tone words from the familiarized language over those from the nonfamiliarized language on 57.43% of test trials (95% CI [54.08, 60.79]), significantly greater than chance, t(29) = 4.34, p < .001, g = 1.11 (see Figure 5). Again, performance did not differ significantly from passive familiarization only (Experiment 1a), t(48) = 1.49, p = .14, g = .42, or active reproduction only (Experiment 1b), t(49) = 1.82, p = .08, g = .51. Furthermore, performance did not differ significantly between the long (Experiment 2a) and short (Experiment 2b) familiarization conditions, t(58) = .59, p = .56, d = -.15, suggesting that longer periods of exposure conferred no additional benefit to statistical learning.

On the reproduction task, data were missing for two sequences, and in the remaining data, there were 137 sequences containing a between-word error (46%), 101 sequences containing a within-word error (34%), 55 sequences containing an error on the first element (18%), and five sequences containing no errors (2%). Participants were more likely to make errors between than within tone words (58:42), differing significantly from chance,  $\chi^2(1) = 86.1$ , p < .001, V = .54 (see Figure 6). There was no difference in error position effects between Experiments 2a and 2b,  $\chi^2(1) = 3.09$ , p = .08, V = .08. The distribution of error positions pooled across Experiments 2a and 2b did not differ from that of Experiment 1b,  $\chi^2(1) = 3.17$ , p = .08, V =.07. These results demonstrated no positive effect of prior passive familiarization-regardless of its duration-on the observed pattern of error positions, over and above the statistical learning that took place online during the reproduction task.

Finally, participants correctly reproduced an average of 8.10 elements (95% CI [7.21, 9.08]), which did not differ from Experiment 2a, t(58) = .98, p = .33, g = .25, suggesting no effect of longer passive exposure on recall. However, Experiment 2a and 2b participants pooled together (M = 8.50, [7.8, 9.2]) unexpectedly reproduced shorter sequences than in Experiment 1b, t(79) = 2.85, p = .006, g = .71. A completely post hoc speculation is that this may have been due to fatigue from the longer passive exposure task and the concurrently presented Raven's matrices test (a possibility we controlled for in Experiment 3, below).

### **Experiment 3**

Contrary to our hypotheses, when passive exposure and active reproduction were combined, or when the duration of passive exposure was increased, we found no enhancement of statistical learning, assessed either by recognition performance, error

### Figure 6

Relative Frequency of Errors Committed Between Tone Words in the Reproduction Task for Experiment 2a (Familiarization + Reproduction) and 2b (Short Familiarization + Reproduction)



*Note.* Chance performance is shown in the rightmost bar; and asterisks indicate performance differing significantly from chance.

position effects, or memory recall. In fact, recall performance was poorer relative to Experiment 1b, perhaps because of the longer and more demanding experimental session.

Experiment 3a controlled for this possibility by combining passive exposure to unstructured sequences, followed by the reproduction task with structured sequences. This achieved a session length equivalent to Experiment 2b without creating additional opportunities for statistical learning. If the poor recall performance in Experiment 2b compared to 1b was due to the longer session length, recall performance in Experiment 3a should be no different from Experiment 2b.

Experiment 3b provided a complementary control condition, combining passive exposure to a structured artificial language with reproduction of unstructured sequences. Here, we would expect error positions to be as predicted by chance. Furthermore, if statistical learning were to influence recall, participants should reproduce shorter sequences than in Experiment 2b, which was identical to Experiment 3b in all respects (including session length) except that it presented structured rather than unstructured sequences in the reproduction task. Recognition performance in Experiments 3a and 3b should be above chance and no different from Experiment 2b unless passive exposure or active reproduction with *unstructured* sequences attenuate the effects of statistical learning.

Experiment 3c provided a baseline control condition, with unstructured sequences used for both passive familiarization and active reproduction. Since there was no opportunity whatsoever for statistical learning, there should be no error position effects in the reproduction task, and performance should be at chance in the recognition task.

### **Participants and Procedure**

Sixty-seven participants (*M* age = 22.63, SD = 4.05, 41 female) completed the experiment. In general, the procedure was identical to that of Experiment 2b (short familiarization + reproduction) except for the following: (a) In Experiment 3a (N = 17, *M* age = 22.41, SD = 3.30, nine female), unstructured sequences were used for passive familiarization (see "General Method"); (b) In Experiment 3b (N = 30, *M* age = 24.5, SD = 4.57, 16 female), unstructured sequences were used for active reproduction; and (c) In Experiment 3c (N = 20, *M* age = 20, SD = 1.78, 16 female), unstructured sequences were used for both passive familiarization and active reproduction. Since the length of exposure did not appear to have an effect on reproduction or recognition in Experiment 2, only a single block of passive familiarization was used in Experiment 3 (as in Experiment 2b: short familiarization + reproduction).

# **Results and Discussion**

Recognition performance for Experiments 3a, b, and c is shown in Figure 7. Error position effects in the reproduction task F7 are shown in Figure 8.

# Experiment 3a (Unstructured Familiarization + Structured Reproduction)

In the recognition task, participants recognized sequences from the artificial language they had encountered in the reproduction task significantly better than chance (M = 56%, 95% CI [50.77, 61.24]), t(16) = 2.25, p = .039, g = .75. Performance was not

# Figure 7





*Note.* Chance performance is at 0.5, shown by the horizontal line, and asterisks indicate performance differing significantly from chance. Error bars represent 95% confidence intervals around the mean.



Relative Frequency of Errors Committed Between Tone Words in the Reproduction Task for Experiments 3a (Unstructured Familiarization + Reproduction), 3b (Familiarization + Unstructured Reproduction), and 3c (Unstructured Familiarization + Unstructured Reproduction)



*Note.* Chance performance is shown in the rightmost bar, and asterisks indicate performance differing significantly from chance.

significantly different from Experiment 2b, where both familiarization and reproduction sequences were structured, t(45) = -.47, p = .64, g = .14.

In the reproduction task, data were missing for one sequence, and in the remaining data, there were 85 sequences containing a between-word error (50%), 46 sequences containing a withinword error (27%), 31 sequences containing an error on the first element (18%), and seven sequences containing no errors (4%). Participants made more errors between tone words than within tone words (65:35), differing significantly from the pattern predicted by chance,  $\chi^2(1) = 75.92$ , p < .001, V = .67, but not from that observed in Experiment 2b,  $\chi^2(1) = 1.60$ , p = .21, V = .07.

Participants reproduced an average of 9.99 elements (95% CI [8.82, 11.17]), which was significantly greater than that achieved in Experiment 2b, t(45) = 2.37, p = .02, g = .71, but no different from that achieved in Experiment 1b, t(36) = .53, p = .60, g = .17. This suggests that session length was not responsible for the relatively poor recall performance in Experiment 2b.

# *Experiment 3b* (*Structured Familiarization* + *Unstructured Reproduction*)

Participants recognized tone words from the familiarized language significantly better than chance (M = 58.54%, 95% CI [55.06, 62.02]), t(29) = 4.81, p < .001, g = 1.23, even after reproducing unstructured sequences. Recognition performance did not differ from Experiment 3a, t(45) = .82, p = .42, g = .24, nor from Experiment 2b, t(58) = .45, p = .65, g = .11.

In the reproduction task, even though the unstructured language was used, we analyzed error position effects in the same way as before, comparing errors made at positions that would correspond to between- and within-word positions had a structured language been used. There were 77 sequences containing a between-word error (26%), 192 sequences containing a within-word error (64%), 27 sequences containing an error on the first element (9%), and four sequences containing no errors (1%). Proportions of errors at between- versus within-word positions (29:71) closely conformed to chance (30:70),  $\chi^2(1) = .24$ , p = .62, V = .03, and differed significantly from those observed in Experiment 3a,  $\chi^2(1) = 46.58$ , p < .001, V = .34, and Experiment 2b,  $\chi^2(1) = 42.17$ , p < .001, V = .29.

Finally, we turn to recall performance. Despite error position effects being at chance unlike in Experiment 2b, average sequence length reproduced (M = 8.67, 95% CI [7.88, 9.46]) did not differ significantly from that in Experiment 2b, t(58) = -.83, p = .41, g = -.21.

# *Experiment 3c (Unstructured Familiarization + Unstructured Reproduction)*

As expected, recognition performance was at chance (M = 50.1%, 95% CI [45.49, 54.72]), t(19) = .04, p = .965, g = .01. It was also significantly less accurate than performance in Experiment 2b, t(48) = -2.58, p = .013, g = .73.

In the reproduction task, there were 52 sequences containing a between-word error (26%), 126 sequences containing a withinword error (63%), 22 sequences containing an error on the first element (11%), and no sequences containing no errors (0%). Proportions of errors at between- versus within-word positions (29:71) did not differ from chance,  $\chi^2(1) = .052$ , p = .82, V = .02, nor from Experiment 3b,  $\chi^2(1) < .001$ , p = .98, V = .001. These results confirm the absence of above-chance recognition performance and error position effects when participants are given no opportunity for statistical learning.

Since it afforded no opportunity whatsoever for statistical learning, Experiment 3c provided a baseline to assess effects of statistical learning on reproduced sequence length. The results were somewhat contradictory, showing that baseline performance (M =8.36, 95% CI [7.66, 9.05]) was poorer than in Experiment 1b (M = 10.4), t(39) = -3.43, p < .001, g = -1.05, and Experiment 3a (M = 9.99), t(35) = -2.43, p = .02, g = -.78, but not significantly different from Experiment 2a (M = 8.85), t(48) = -.69, p = .50, g =-.19, or Experiment 2b (M = 8.1), t(48) = .32, p = .75, g = .09. This raises the question of cohort effects producing differences between experiments that exceed any potential effects of statistical learning. To address this, we pooled data from those experiments (1b, 2a, 2b, and 3a) in which there were opportunities for statistical learning to influence recall. The cross-experiment mean sequence length (M = 9.16, [8.63, 9.7]) did not differ significantly from the baseline provided by Experiment 3c, t(116) = -1.29, p = .20, g =-.31. Indeed, the difference in mean performance was less than a single sequence element. These results suggest that statistical learning-and the apparent emergence of chunking evidenced by the error position effects-has very little consistent and reliable impact on recall performance.

# **Experiment 4**

The results of Experiments 1, 2, and 3 demonstrate that error position effects provide a robust measure of dynamic statistical learning during the Simon reproduction task. However, these effects have been demonstrated only for sequences constructed from a particular set of pure tones, which were always accompanied by a corresponding and simultaneously occurring visual sequence in that the colored response button corresponding to each tone flashed when that tone was played. Above-chance recognition performance in Experiments 2 and 3a demonstrates auditory learning, but it remains possible that an independent, parallel process of visual statistical learning of color sequences was responsible for the error position effect.

To unambiguously establish that the error position effect resulted purely from auditory statistical learning and extended beyond the pure tones used in the previous experiments, Experiment 4 used an altered Simon reproduction task where we replaced the tones with four environmental sounds and eliminated the flashing colored lights that previously corresponded to each tone. This was intended to ensure that participants would encode and remember the sequence of sounds presented to them before translating them into a sequence of movements to press the corresponding buttons during recall (these movements would be visually guided initially but could become proprioceptively guided throughout the task).

This experiment also extended the paradigm to formally trained musician participants so as to assess whether the error position measure of statistical learning was influenced by musical training. Since Tierney et al. (2008) found that musicians were able to reproduce longer auditory sequences (spoken color names) than nonmusicians (but cf. Carey et al., 2015), we hypothesized that musicians would show better recall performance in the reproduction task than nonmusicians. However, since musicians do not appear to show better statistical learning ability than nonmusicians in artificial grammar learning studies (Loui et al., 2010; Rohrmeier et al., 2011), we hypothesized that error position effects would not differ between the groups.

# **Participants and Procedure**

Nineteen nonmusicians (*M* age = 26.5, SD = 2.5, nine female) and 20 musicians (*M* age = 22.8, SD = 2.9, 13 female) with no reported history of hearing, visual, or neurological impairments completed the experiment after giving informed consent; all were paid for their participation. Nonmusicians were recruited as before, except that all had completed or were enrolled in a university master's or doctorate degree to approximate the education level of the musician group. Musicians were recruited from conservatories in London; all but one was completing, or had completed, a performance degree. All had at least 10 years of continuous formal musical training (*M* 14.0 years, SD = 3.3 year) on piano, guitar, voice, recorder, or orchestral instruments and began music practice on average at the age of 6.5 years (SD = 1.9 years).

In general, the procedure was identical to that of Experiment 1b (reproduction only) with the following differences. Participants completed the Simon reproduction task for 20 sequences, 10 of which included tones and synchronously timed colored lights as before, while the other 10 used environmental sounds with no colored lights associated with each sound. For environmental-sound stimuli, the interface presented four colored buttons (red, blue, green, and yellow) as for tone sequences but added a picture of the corresponding sound superimposed on each button. Crucially, and unlike the tone condition, the colored buttons did not flash during presentation of the environmental sound corresponding to that button.

Tone sequences and environmental-sound sequences were presented five at a time and were interleaved. The same language was used to generate the sequences of tones and environmental sounds, but distinct sequences were used for the two conditions. Participants were randomly assigned to one of two presentation orders such that one group completed five tone sequences first, while the other group completed five environmental-sound sequences first. As in Experiment 1b, there was no initial stage of passive familiarization; after completing the reproduction task, participants performed a 2AFC recognition task, in a shorter, 32-trial format with 16 trials each for environmental-sound and pure-tone stimuli.

The environmental-sound stimuli were four short environmental sounds taken from Leech et al. (2009): a camera shutter, a drink being poured, a phone being dialed, and a doorbell. These were chosen as they were short and easy to recognize, as shown by Krishnan et al. (2013) who found that both children and adults had high mean accuracy scores for these four sounds in an environmental-sound identification task. The duration of each original sound was shortened to 300 ms using Adobe Audition such that they had the same length as the pure tones used in the previous experiments. Analogous to the tone experiments, each button was paired with a fixed 300-ms environmental sound (red button, phone dial; blue button, drink pour; green button, camera shutter; yellow button, doorbell).

The experiment was conducted in a sound-attenuated room and presented using an Apple 13-in. Macbook Pro laptop with Psychophysics Toolbox Version 3 (Brainard, 1997) in MATLAB (2010a; 64 bit). Auditory stimuli were presented through Sennheiser HD- 380 Pro headphones, via an ESI UGM 96 24-bit external sound card, connected to the laptop by USB. Participant responses were collected using a Logitech Precision gamepad in the Simon sequence reproduction tasks, and the laptop keyboard was used for the 2AFC recognition task. Participants held the gamepad while sitting in front of the laptop, which was placed on a desk.

# **Results and Discussion**

In the recognition task, performance was above chance for puretone stimuli (M = 63.50%, 95% CI [58.10%, 68.92%]), t(38) =5.05, p < .001, g = .81, and environmental-sound stimuli (M =57.30%, [51.68%, 62.92%]), t(38) = 2.63, p = .012, g = .42 (see Figure 9). Musicians and nonmusicians did not differ significantly in their recognition performance for either pure-tone stimuli, t(37) =1.24, p = .22, g = .39, or environmental-sound stimuli, t(37) =.55, p = .59, g = .17.

Three musicians were removed from the error position analyses due to technical difficulties accessing their data; thus, error position analyses are from 19 nonmusicians and 17 musicians. Error position data are shown in Figure 10.

For the pure-tone reproduction task, there were 164 sequences containing a between-word error (46%), 129 sequences containing a within-word error (36%), 29 sequences containing an error on the first element (19%), and no error-free sequences (0%). The proportion of errors made between relative to within words (56:44) was significantly different from that predicted by chance (30:70),  $\chi^2(1) = 94.12$ , p < .001, V = .51.

For the environmental-sound reproduction task, there were 145 sequences containing a between-word error (41%), 149 sequences containing a within-word error (42%), 62 sequences containing an error on the first element (17%), and no error-free sequences (0%).

The proportion of errors made between relative to within words (49:51) was significantly different from that predicted by chance (30:70),  $\chi^2(1) = 52.26$ , p < .001, V = .38, but not significantly different from the error position effect observed for tone sequences,  $\chi^2(1) = 2.35$ , p = .13, V = .06. Furthermore, the error position effect did not differ between musicians and nonmusicians for tone sequences,  $\chi^2(1) = .008$ , p = .93, V = .005, or for environmental-sound sequences,  $\chi^2(1) = .31$ , p = .58, V = .03. The error position effect was observed consistently across levels of musical training for both pure-tone sequences and environmental-sound sequences. Furthermore, the effect was not dependent on the particular pure tones used in Experiments 1, 2, and 3, nor on the sequence of colored lights presented alongside tone sequences. Rather, the effect extended to statistically structured sequences of natural sounds presented without accompanying colors.

Overall, participants correctly reproduced an average of 11.04 elements (95% CI [9.97, 12.11]) in the tone condition and 9.20 elements ([8.07, 10.34]) in the environmental-sounds condition. A mixed-design analysis of variance with sound type (tone, environmental sound) as a within-subject variable and group (musician, nonmusician) as a between-subjects variable showed that participants reproduced significantly longer tone versus environmentalsound sequences (Greenhouse-Geisser corrected), F(1, 37) =26.46, p < .001. Musicians also reproduced significantly longer sequences overall than nonmusicians, F(1, 37) = 16.24, p < .001, with an average of 11.83 elements reproduced ([10.43, 13.22]) compared with an average of 8.32 reproduced by nonmusicians ([7.16, 9.49]). There was no interaction between sound type and group, F(1, 37) = 1.16, p = .288. The better recall performance for musicians than nonmusicians for both tone and natural sound sequences may reflect differences in cognitive abilities such as working memory rather than statistical learning ability given that

### Figure 9





*Note.* Chance performance is at 0.5, shown by the horizontal line, and asterisks indicate performance differing significantly from chance. Error bars represent 95% confidence intervals around the mean.

Relative Frequency of Errors Committed Between Words in the Reproduction Task for Experiment 4 With Environmental Sounds (Leftmost Bar) and Pure Tones (Middle Bar)



*Note.* Chance performance is shown in the rightmost bar, and asterisks indicate performance differing significantly from chance.

no differences were observed between the groups in either recognition performance or error position effects as measures of statistical learning.

# **Experiment 5**

There is one remaining factor that may have influenced the error position effect observed in the preceding experiments, namely the consistent pairing of tones/sounds and buttons across participants. The advantage of using a single tone-button assignment for all participants is that it minimizes potential item effects on estimates of cross-participant variability and therefore increases sensitivity to potential individual differences (see below) or group effects (see Experiment 4). However, it is possible that the error position effects could be driven in part by the specific spatial configuration that was chosen, to the extent that the tone words in both languages correspond to spatially contiguous buttons. Specifically, the elements were mapped to buttons such that all the responses within a "tone word" would go in either a clockwise or counterclockwise direction. Therefore, increased errors between versus within words could also be a consequence of visuomotor factors.

Experiment 5 addressed this question by fully randomizing tone-to-button assignment in the Simon reproduction task across both structured artificial languages. We predicted that error position effects would still be observed when there was no consistent spatial mapping of tones to buttons across participants.

### **Participants and Procedure**

Forty-eight participants (M age = 24.06, SD = 5.65, 18 female) completed the experiment. In general, the procedure was identical to that of Experiment 1b (reproduction only) with the following

differences: (a) Participants were recruited online via Prolific and undertook a web-based version of the Simon sequence reproduction task; (b) There was no prior familiarization phase, nor any subsequent recognition task; and (c) For each language, each participant was assigned to one of the 24 possible unique assignments of tone frequency to colored button.

#### **Results and Discussion**

On the reproduction task, there were 214 sequences containing a between-word error (45%), 162 sequences containing a withinword error (34%), 104 sequences containing an error on the first element (22%), and no sequences containing no errors (0%). The proportion of errors made between relative to within tone words (57:43) was significantly different from that predicted by chance  $(30.70), \chi^2(1) = 129.7, p < .001, V = .52$  (see Figure 11). It was also significantly different from the proportions observed in experiments with unstructured reproduction: Experiment 3b (structured familiarization + unstructured reproduction),  $\chi^2(1) = 49.55$ , p < .001, V = .28; Experiment 3c (unstructured familiarization + unstructured reproduction),  $\chi^{2}(1) = 36.04$ , p < .001, V = .26. By contrast, the error position effect in the present experiment was not significantly different from that observed in any of the experiments with structured reproduction: Experiment 1b (reproduction only),  $\chi^2(1) = .89$ , p = .34, V = .04; Experiment 2a (familiarization + reproduction),  $\chi^{2}(1) = 3.28$ , p = .07, V = .07; Experiment 2b (short familiarization + reproduction),  $\chi^2(1) = .01$ , p = .94, V =.003; Experiment 3a (unstructured familiarization + structured reproduction),  $\chi^2(1) = 2.23$ , p = .14, V = .07.

Finally, we turn to recall performance. Participants correctly reproduced an average of 6.99 elements (95% CI [6.46, 7.52]).

Relative Frequency of Errors Committed Between Words in Experiment 5, Where Each Participant Experienced a Different Assignment of Tone Frequency to Button Position



*Note.* Chance performance is shown in the rightmost bar, and asterisks indicate performance differing significantly from chance.

These results confirm that the error position effect persisted when the spatial position of the buttons associated with each tone was randomly assigned across participants, thus demonstrating that it was not dependent on a particular spatial configuration of the response buttons. The error position effect was also not significantly weaker than that observed in the previous experiments with structured sequence reproduction (Experiments 1b, 2a, 2b, and 3a).

### **Pooled Analyses Across Experiments**

Pooling data across experiments provides sufficient power to address two questions of interest. First, can the error position effect illustrate the trajectory of learning across the reproduction task? Second, does performance on the reproduction task correlate with performance on the recognition task? The analyses reported below are Pearson correlations (except where noted), with Bayes factors (BF) also computed.

# **Error Position Effect Across Sequences**

The error position effect observed for structured sequences in Experiments 1b, 3a, 4, and 5 (and not observed for unstructured sequences in Experiments 3b and 3c) must result from online statistical learning during the reproduction task. To examine in more detail the time course of this learning, we conducted a correlation

analysis of how the error position effect changed across the 10 sequences making up the active reproduction task, pooled across multiple experiments for the structured languages and, separately, the unstructured language. The results showed a significant linear increase in proportion of between-word errors for structured sequences (Experiments 1b, 2, 3a, 4, and 5), r(8) = .83, p < .01, BF = 7.41, but no significant change for unstructured sequences (Experiments 3b and 3c), r(8) = -.1, p = .79, BF = .63. Siegelman, Bogaerts, Kronenfeld, and Frost (2018) found that learning followed a logarithmic rather than a linear trajectory in their RTbased measure of visual statistical learning. The same is true for structured sequences in the present experiments where a logarithmic trajectory, r(8) = .91, p < .01, BF = 27.23, shown in Figure 12, provided a better fit than a linear trajectory, BF = 7.93. These results confirm the dynamics of learning across the reproduction task, suggesting a logarithmic trajectory of learning.

Across all experiments with structured sequences in the reproduction task (Experiments 1b, 2, 3a, 4, and 5), the error position effect on the 10th and final sequence (65:35) differed significantly from that expected by chance (30:70),  $\chi^2(1) = 91.42$ , p < .001, V = .76, while this was not the case for the first sequence (33:67),  $\chi^2(1) = .72$ , p = .40, V = .06. Furthermore, focusing specifically on Experiment 2 where the reproduction task was preceded by passive exposure to structured sequences, a significant error position effect was still observed for the 10th sequence (51:49),  $\chi^2(1) =$ 9.9, p < .01, V = .46, but not for the first sequence (25:75),  $\chi^2(1) =$ .59, p = .44, V = .10. These results suggest very little if any transfer of learning between passive exposure and active reproduction.

# Correlation Between Reproduction and Recognition Performance

Pooling data from all experiments that contained the reproduction task with structured sequences and the recognition task (Experiments 1b, 2, 3a, and 4) allows an analysis of whether performance on these tasks is correlated across participants. In other words, is it the case that participants who show strong statistical learning ability on the reproduction task also show strong statistical learning assessed by the recognition task? We would expect this to be the case if both tasks were probing the same underlying psychological mechanisms of statistical learning. However, there was no significant correlation (Spearman rank given that sequence length was not normally distributed) between recognition performance (percent correct) and either mean sequence length, r(168) =.14, p = .08, BF = 1.05, or proportion of between-word errors, r (168) = -.01, p = .85, BF = .18. This suggests that the recognition and reproduction tasks reflect at least partially disjoint underlying psychological representations or processes related to statistical learning. In other words, participants whose error positions showed evidence of greater statistical learning (i.e., those making a greater proportion of errors between rather than within tone words) were no more likely to achieve above-chance performance on the recognition task.

Pooling across the same set of experiments, there was a significant correlation between the proportion of between-word errors a participant made on the reproduction task and their mean sequence length, r(168) = .39, p < .001, BF = 38,245. This shows that participants with strong statistical learning, reflected by the error position effect, also recalled longer sequences, suggesting an influence of statistical

Change in the Proportion of Between-Word Errors Across the 10 Sequences Presented in the Reproduction Task



*Note.* Filled circles plot the data for experiments with structured sequences (Experiments 1b, 2, 3a, 4, and 5), with a logarithmic regression line fitted, while unfilled circles are for experiments with unstructured sequences (Experiments 3b and 3c), with a linear regression line fitted. The shaded areas represent 95% confidence intervals around the fitted regression lines.

learning on recall. However, it is also possible that one or more general cognitive abilities, such as working memory or auditory attention, influenced both statistical learning and recall performance. Furthermore, the absence of a significant correlation in Experiment 5, r(46) = .17, p = .27, BF = .56, raises the possibility that the significant correlation in Experiments 1–4 could reflect differences in experimental design including the spatial configuration (which may have improved recall performance for some participants in Experiments 1–4 and, correspondingly, impaired recall performance in Experiment 5) or the slower initial timing used in the earlier experiments, which may have facilitated recall and overall task performance. However, this could also represent a Type-II error given the poorer and relatively homogenous recall performance in Experiment 5. This warrants further examination in future research.

### **General Discussion**

Overall, the results of these five experiments demonstrate that the error position effect during sequence reproduction provides a robust, online processing-based measure of statistical learning that addresses many of the limitations of the reflection-based 2AFC recognition task predominantly used to assess statistical learning (Christiansen, 2019; Frost et al., 2019; Isbilen et al., 2020; Siegelman, Bogaerts, Christiansen, & Frost, 2017). The error position effect during reproduction has clear advantages as an estimate of implicit statistical learning in terms of errors committed on individual elements making up auditory sequences, which can be captured dynamically throughout the period that statistical learning is actually taking place. Indeed, the results

showed that the error position effect strengthens throughout the reproduction task as participants gained greater experience of the statistical regularities underlying the sequences they were reproducing. The error position effect extended naturally to sequences of environmental sounds without accompanying visual cues and across levels of musical training. By contrast, recall performance showed significant intergroup variability and failed to clearly distinguish conditions providing an opportunity for statistical learning from those that did not. Somewhat surprisingly, the error position effect showed little influence of prior passive exposure to sequences with the same statistical structure, suggesting that knowledge acquired from passive exposure did not transfer directly to active reproduction. Furthermore, while recognition performance provided robust evidence of statistical learning, it did not vary with the amount of exposure provided through passive familiarization, active reproduction, or both. Finally, as hypothesized based on previous findings with other processingbased measures of auditory statistical learning (Batterink et al., 2015; Franco et al., 2015), the error position effect was uncorrelated with performance on the 2AFC recognition task, suggesting that these tasks reflect at least partially distinct underlying psychological abilities. In the following, we discuss these results and their underlying psychological representations and processes in relation to other findings in the literature.

# The Error Position Effect: A New Processing-Based Measure of Statistical Learning

Although it has dominated the literature on statistical learning, the 2AFC recognition task has been criticized for providing an offline, binary, and highly indirect measure of statistical learning (Christiansen, 2019; Frost et al., 2019; Isbilen et al., 2020; Siegelman et al., 2017). Like other processing-based measures of statistical learning, such as the SRT task, the error position measure has three primary advantages over traditional reflection-based measures, such as the 2AFC recognition task. First, it is recorded during the time that statistical learning is taking place, rather than afterward. Second, it has greater granularity, allowing for analysis of the trajectory of statistical learning. Third, it does not require additional psychological processes related to explicit decision-making, making for a purer measure of statistical learning. The error position measure revealed a logarithmic trajectory of learning across the task, beginning with a steep profile that progressively flattened with increasing experience. This contrasts with other possible learning profiles, including linear or step functions, and is consistent with the trajectory of learning reported by Siegelman, Bogaerts, Kronenfeld, and Frost (2018).

The error position metric of statistical learning clearly distinguished the experiments that provided an opportunity for statistical learning from those that did not and appeared to do so consistently across different experimental variants. It was observed for sequences of both pure tones and environmental sounds, suggesting that it is not specific to one particular kind of auditory stimulus but rather extends to statistical structure across auditory domains. Finally, the error position effect is robust, showing similar effects between the various participant groups used in the different experiments. In particular, in Experiment 4, musicians showed error position effects that were no different from those of nonmusicians, consistent with previous evidence that musicians show no better artificial grammar learning performance than nonmusicians (Loui et al., 2010; Rohrmeier et al., 2011).

# The Relationship Between Passive Exposure and Active Reproduction

One of the most surprising findings in the present research is the lack of any additive effect of statistical learning across passive exposure and active reproduction tasks. Although performance on the recognition task was above chance and comparable with previous research (e.g., Saffran et al., 1999), it showed no significant difference between conditions regardless of how much passive familiarization and/or active reproduction was undertaken. This might be taken to indicate a performance ceiling—but the distribution of the data does not show ceiling effects. Rather, we suggest that above-chance performance on the sequence recognition task depends only on some exposure to the statistically structured sequences regardless of whether this involves passive exposure, active reproduction, or a combination of the two.

As a result, it is difficult to draw general comparisons about how strongly passive exposure and active reproduction engage statistical learning mechanisms. However, it is striking that equivalent overall recognition performance resulted from passive exposure to 2,070 tones in Experiment 1a and experience of a median of 1,382 tones in Experiment 1b (including both presentation and reproduction). Although one should not overgeneralize from relatively short periods of learning, it is also noteworthy that while some participants in the reproduction task heard over 10 times as many tone words as other participants (due to their success in reproducing much longer sequences), this differential experience did not significantly affect recognition accuracy. The relative strength with which different tasks engage statistical learning could be investigated directly in future research by systematically varying the length of the sequences presented in passive exposure and active reproduction. One caveat is that exposure to sequences in the reproduction task differed in (at least) one additional way from passive listening. Sequences in the reproduction task always started with a whole tone word. Therefore, across all 10 sequences reproduced, the onset of most or all of the four tone words would be associated with the onset of the sequence. This may provide an "edge cue" for segmenting word onsets. This could be addressed by presenting multiple shorter sequences in passive exposure, rather than long continuous sequences.

It was surprising to find that the trajectory of statistical learning during the reproduction task started from a level very close to chance performance, suggesting little, if any, carryover of statistical learning from passive exposure. This raises the possibility that knowledge acquired through passive exposure may have (partially) disjoint storage in memory from that acquired through active reproduction. If this were the case, it could also account for the lack of additive effects of passive exposure and active reproduction. One possibility (discussed further below) is that the reproduction task leads to encoding of sensorimotor representations of statistical learning, distinct from the purely auditory representations encoded during passive exposure.

# The Relationship Between Reproduction and Recognition Performance

The divergence between the reproduction and recognition tasks in the present experiments might be taken as evidence that the two tasks tap different underlying knowledge representations, one unconscious (often termed "implicit") and the other conscious, declarative, or "explicit" (Batterink et al., 2015)-for a critical review of this dichotomy, see Shanks and St. John (1994). Following Franco et al. (2015), however, a different-though potentially complementary-explanation points rather to the retrieval processes recruited by the two tasks (see also Shanks & Perruchet, 2002). Error positions in the reproduction task reflect a process of recall for entire sequences, extended by one element on each trial, which does not necessarily require the sequence to be segmented or chunked but rather may rely on sensitivity to the statistical structure of the continuous sequences as they are processed. By contrast, the recognition task presumably requires segmentation and retrieval of three-element words from stored representations of the sequence for matching to the words presented as isolated units in the 2AFC task.

In other words, it is possible that the error position effect reflects representations in memory of statistical information in the form of conditional probabilities, perhaps encouraged by the nature of the task in which a single element is added on each trial. Conversely, the recognition task may encourage the formation of chunk-based representations in which the three-element words are segmented. If the two tasks involve different kinds of representation in memory, then it is plausible that participants could vary independently in their capability for encoding and processing each type of representation. Consistent with this account is evidence that SRNs have accurately simulated processing-based measures of statistical learning tasks such as the SRT (Franco & Destrebecqz, 2012; Misyak et al., 2010b) and HRE (Rey et al., 2020), whereas the chunking model PARSER has successfully modeled 2AFC recognition performance (Perruchet & Vinter, 1998). Furthermore, Siegelman et al. (2019) ran computational simulations of statistical learning in a self-paced visual task (Siegelman, Bogaerts, Kronenfeld, & Frost, 2018), finding evidence of substantial variance in the extent to which individuals prioritize local co-occurrences of elements (providing the basis for conditional probability estimation) on the one hand or global patterns (providing the basis for chunking) on the other. The present experiments were not designed to distinguish between these theoretical accounts, but the initial results and general paradigm provide a fruitful avenue for future investigation.

# No Effect of Statistical Learning on Recall Performance

The present results also allowed us to assess another processing-based measure of statistical learning-recall performance reflected by the average length of the sequences reproduced on the active reproduction task-and compare it to the error position measure. Recall performance has been successfully used as a measure of statistical learning in previous research with syllable sequences (Conway et al., 2010; Isbilen et al., 2020; Karpicke & Pisoni, 2004). However, in contrast to the error position effects, we found no evidence of greater recall performance in those experiments that provided an opportunity for statistical learning, compared with experiments that used random sequences and therefore provided no opportunity for statistical learning. A plausible explanation is that this reflects large variance in recall performance between participants that swamps any effects of statistical learning that might be present. This is especially apparent in Experiment 4, where musicians showed better recall performance (longer mean sequence length) than nonmusicians, even though there was no difference in error position effects or recognition performance between the groups, thus suggesting that statistical learning ability per se did not differ, consistent with previous findings (Loui et al., 2010; Rohrmeier et al., 2011). It is possible that this result reflects better working and STM in our musician group, which would also be consistent with previous results (Franklin et al., 2008). On the other hand, using a paradigm very similar to the active reproduction task employed in the present experiments, Carey et al. (2015) did not find any difference in recall performance between nonmusicians and a group of musicians with at least as much musical training as the musicians who participated in Experiment 4. In light of these considerations, we suggest that recall performance may constitute a less robust and precise processing-based measure of statistical learning in reproduction tasks than the error position effect.

It is somewhat surprising that between-experiment differences in average sequence length reproduced did not appear to reflect how much passive familiarization and/or active reproduction were undertaken. We had hypothesized that learning of statistical regularities in the structured artificial languages would allow the sequences presented during the reproduction task to be segmented and stored as chunks in memory. Given that participants can recall longer sequences when they can be segmented into chunks (Miller, 1956), we expected this would lead to reproduction of longer sequences. However, there was little evidence that participants consistently reproduced longer sequences in experiments where there was an opportunity for statistical learning, compared to experiments where there was not.

One interpretation alluded to above is that, while participants learned transitional probabilities during exposure and reproduction, they did not actually store the sequences as chunks in memory—or at least if they did, the chunks formed did not coincide with the statistical structure of the sequences. This would be inconsistent with recent theoretical proposals that statistical learning relies on low-level memory mechanisms such as chunking (Christiansen, 2019; Perruchet, 2019; Thiessen, 2017) but consistent with an alternative proposal that probability estimation and chunking operate as parallel, interconnected psychological processes (Isbilen et al., 2020), which could be emphasized to different extents by different tasks.

A recent study by Isbilen et al. (2020) provides evidence that implicit statistical learning can have positive effects on recall performance for syllable sequences (see also Conway et al., 2010). Participants were passively exposed to artificial syllable sequences created by pseudorandomly concatenating six trisyllabic nonsense words (cf. Saffran, Aslin, & Newport, 1996). This was followed by an oral recall task for valid and invalid word pairs, with invalid pairs created by reordering the syllables to remove transitional probability information that would allow for identification of word boundaries. In contrast to the present findings, the results showed better recall for valid compared to invalid stimuli. However, since Isbilen et al. did not include a condition where the recall task was not preceded by passive familiarization (as in our Experiments 1b and 3a), it is impossible to say whether their results reflect learning during passive familiarization or dynamically acquired during the recall task.

The inconsistency between the present results and those of Isbilen et al. (2020) may be due to differences in familiarity between the stimulus domains. It is possible that orally reproducing sequences of familiar syllables engaged phonological (or auditory-motor) chunking mechanisms more readily than the abstract sound (and color) sequences used in the present research. Some support for this interpretation can be gleaned from the slightly better recognition performance in their study compared to ours (68% and 76% in their first and second sessions vs. 62% in Experiment 2). More significantly, Isbilen et al. (2020) found that recall was significantly predicted by natural language statistics (bigram and trigram frequency in large corpora of spoken English), suggesting an influence of prior language learning. Similarly, Siegelman, Bogaerts, Elazar, et al. (2018) have shown that statistical learning of sequences of syllables, but not sequences of visual or nonverbal auditory stimuli, exhibits item-specific effects resulting in low correlations between items that were related to natural language statistics. Additional support for a difference between phonological/oromotor and pitch-motor encoding comes from a recent study showing no correlation, across participants, in statistical learning ability between verbal and nonverbal stimuli within modality (Siegelman & Frost, 2015).

Therefore, it is possible that the abstract materials used in the present research engaged chunking mechanisms less strongly than the more familiar linguistic materials used by Isbilen et al. (2020) and Conway et al. (2010). Further research is required to assess any potential influence of domain familiarity on the present results. More generally, the results should be replicated across different modalities given evidence of modality-specific differences

in statistical learning (Conway & Christiansen, 2005, 2006; Frost et al., 2015; Tierney et al., 2008).

# Auditory, Visual, Visuomotor, or Auditory-Motor Learning?

Given that the reproduction task is multimodal, it is worth examining in more detail the nature of the learned representations. Above-chance auditory recognition performance in Experiments 1b, 3a, and 4 provides convincing evidence of auditory statistical learning during the reproduction task. Furthermore, the results of Experiment 4 would appear to rule out purely visual statistical learning as an explanation for the error position effect. However, even in Experiment 4, there is still a visuomotor component to the reproduction task because participants must press the button corresponding to each sound in the memorized sequence. We can consider three possible scenarios for the memory representation underlying the error position effect.

First, participants in Experiment 4 may have encoded and memorized the sequence of environmental sounds as it was presented and then, during recall of the sequence, translated each sound into a motor command. Second, it is possible that during presentation of the sequence, each sound was immediately translated into a motor command, and the resulting sequence of motor commands was encoded in parallel with the sequence of sounds, with the error position effect reflecting statistical learning of the motor sequence rather than the auditory sequence. Third, it is possible that the sequences were encoded as a combined multimodal auditorymotor sequence. Unlike a unimodal auditory representation (the first scenario), an auditory-motor memory representation (the third scenario) has the advantage of accounting neatly for the lack of correlation between performance on the reproduction and recognition tasks and the lack of transfer from passive exposure to active reproduction. It is also more parsimonious than the second scenario as it is not absolutely necessary to encode the motor sequence to perform the task.

Future research should investigate these possibilities further. One option would be to conduct experiments with different response modalities in the reproduction task, such as singing. However, this still involves a sequence of motor commands, and in fact, it is difficult to conceive of a reproduction task without any motor component at all. Another complementary possibility would be to introduce a motor component into passive exposure and the recognition task such that each element of the sequences presented elicits a response from participants. If this encourages formation of auditory-motor representations, we might expect increased transfer of knowledge from passive exposure to active reproduction or better correspondence between reproduction and recognition performance. It would also be interesting to use different combinations of visual and auditory sequences in the three tasks making up the experiment (passive exposure, reproduction, and recognition) to investigate transfer of stored representations between tasks when the modality changes. In this context, we would predict that the error position effect would extend naturally to experiments on visual statistical learning (Conway et al., 2010; Fiser & Aslin, 2002; Karpicke & Pisoni, 2004; Kirkham et al., 2002), but transfer between auditory and visual modalities remains to be investigated.

# Conclusion

To conclude, the present results paint a picture of different psychological representations and processes being involved in different tasks related to statistical learning. First, passive exposure to statistically structured auditory sequences may lead to auditory representations in memory of sequential statistical regularities. Second, by contrast, reproduction tasks may lead to auditorymotor representations of sequential statistical regularities in the input. Third, the 2AFC recognition task requires retrieval of stored sequence representations as chunks, segmented according to the statistical regularities and therefore requiring psychological processes additional to those required for learning statistical regularities in the reproduction task. In doing so, the present research introduced a new processing-based measure of statistical learning based on reproduction error positions that shows sensitivity to acquired sequential structure in the absence of clear effects on recall. Furthermore, individual differences in error position effects bear no relationship with individual differences in recognition performance, a prominent reflection-based measure of implicit statistical learning, suggesting that these two measures may reflect distinct underlying psychological processes. The error position metric has advantages over recognition performance: First, it does not require explicit judgements and thus potentially reflects more closely what is thought to be an implicit process of knowledge acquisition. Second, sequence reproduction errors are reflections of the moment-to-moment process during which participants are acquiring knowledge of sequential regularities in a domain. By contrast, recognition judgments reflect what statistical information can be remembered "after the fact."

Though we have focused here on behavioral measures, there is growing understanding of the neural basis of statistical learning (Conway, 2020; Williams, 2020), and since neural measures obviate the need for an overt response, they can potentially provide a more sensitive measure of implicit processes involved in statistical learning (Batterink et al., 2019). Indeed, dynamic changes relating to implicit or incidental learning and knowledge acquisition have been measured using neuroimaging (e.g., Itthipuripat et al., 2017; Lim et al., 2019), and the error position measure may provide a useful behavioral counterpart to such dynamic neural indicators of statistical learning (e.g., Batterink & Paller, 2017). The present results suggest that error position effects hold great promise for helping to understand the dynamic process of acquiring knowledge during statistical learning under various task conditions. This warrants their use across a wider range of reproduction tasks and stimuli varying in modality, domain, and type of statistical structure.

### References

- 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
- Bertels, J., Franco, A., & Destrebecqz, A. (2012). How implicit is visual statistical learning? Journal of Experimental Psychology: Learning,

Memory, and Cognition, 38(5), 1425-1431. https://doi.org/10.1037/ a0027210

- Bogaerts, L., Siegelman, N., Ben-Porat, T., & Frost, R. (2018). Is the Hebb repetition task a reliable measure of individual differences in sequence learning? *The Quarterly Journal of Experimental Psychology*, *71*, 892–905. https://doi.org/10.1080/17470218.2017.1307432
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436. https://doi.org/10.1163/156856897X00357
- Carey, D., Rosen, S., Krishnan, S., Pearce, M. T., Shepherd, A., Aydelott, J., & Dick, F. (2015). Generality and specificity in the effects of musical expertise on perception and cognition. *Cognition*, 137, 81–105. https:// doi.org/10.1016/j.cognition.2014.12.005
- Christiansen, M. H. (2019). Implicit statistical learning: A tale of two literatures. *Topics in Cognitive Science*, 11(3), 468–481. https://doi.org/10 .1111/tops.12332
- Christiansen, M. H., & Chater, N. (2016). Creating language: Integrating evolution, acquisition, and processing. MIT Press. https://doi.org/10 .7551/mitpress/10406.001.0001
- Cleary, M., Pisoni, D. B., & Geers, A. E. (2001). Some measures of verbal and spatial working memory in eight- and nine-year-old hearingimpaired children with cochlear implants. *Ear and Hearing*, 22(5), 395–411. https://doi.org/10.1097/00003446-200110000-00004
- Conway, C. M. (2020). How does the brain learn environmental structure? Ten core principles for understanding the neurocognitive mechanisms of statistical learning. *Neuroscience and Biobehavioral Reviews*, 112, 279–299. https://doi.org/10.1016/j.neubiorev.2020.01.032
- Conway, C. M., Bauernschmidt, A., Huang, S. S., & Pisoni, D. B. (2010). Implicit statistical learning in language processing: Word predictability is the key. *Cognition*, 114(3), 356–371. https://doi.org/10.1016/j .cognition.2009.10.009
- Conway, C. M., & Christiansen, M. H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(1), 24–39. https://doi.org/10.1037/0278-7393.31.1.24
- Conway, C. M., & Christiansen, M. H. (2006). Statistical learning within and between modalities: Pitting abstract against stimulus-specific representations. *Psychological Science*, *17*(10), 905–912. https://doi.org/10 .1111/j.1467-9280.2006.01801.x
- Creel, S. C., Newport, E. L., & Aslin, R. N. (2004). Distant melodies: Statistical learning of nonadjacent dependencies in tone sequences. *Journal* of Experimental Psychology: Learning, Memory, and Cognition, 30(5), 1119–1130. https://doi.org/10.1037/0278-7393.30.5.1119
- Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, *14*(2), 179–211. https://doi.org/10.1207/s15516709cog1402\_1
- Emberson, L. L., Conway, C. M., & Christiansen, M. H. (2011). Timing is everything: Changes in presentation rate have opposite effects on auditory and visual implicit statistical learning. *The Quarterly Journal of Experimental Psychology*, 64(5), 1021–1040. https://doi.org/10.1080/17470218 .2010.538972
- Fiser, J., & Aslin, R. N. (2002). Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(3), 458–467. https://doi .org/10.1037/0278-7393.28.3.458
- Franco, A., & Destrebecqz, A. (2012). Chunking or not chunking? How do we find words in artificial language learning? Advances in Cognitive Psychology, 8(2), 144–154. https://doi.org/10.2478/v10053-008-0111-3
- Franco, A., Eberlen, J., Destrebecqz, A., Cleeremans, A., & Bertels, J. (2015). Rapid serial auditory presentation: A new measure of statistical learning in speech segmentation. *Experimental Psychology*, 62(5), 346–351. https://doi.org/10.1027/1618-3169/a000295
- Franklin, M. S., Sledge Moore, K., Yip, C.-Y., Jonides, J., Rattray, K., & Moher, J. (2008). The effects of musical training on verbal memory. *Psychology of Music*, *36*(3), 353–365. https://doi.org/10.1177/0305735607086044

- French, R. M., Addyman, C., & Mareschal, D. (2011). TRACX: A recognition-based connectionist framework for sequence segmentation and chunk extraction. *Psychological Review*, 118(4), 614–636. https://doi .org/10.1037/a0025255
- Frost, R., Armstrong, B. C., & Christiansen, M. H. (2019). Statistical learning research: A critical review and possible new directions. *Psychological Bulletin*, 145(12), 1128–1153. https://doi.org/10.1037/bul0000210
- Frost, R., Armstrong, B. C., Siegelman, N., & Christiansen, M. H. (2015). Domain generality versus modality specificity: The paradox of statistical learning. *Trends in Cognitive Sciences*, 19(3), 117–125. https://doi.org/ 10.1016/j.tics.2014.12.010
- Hay, J. F., Pelucchi, B., Graf Estes, K., & Saffran, J. R. (2011). Linking sounds to meanings: Infant statistical learning in a natural language. *Cognitive Psychology*, 63(2), 93–106. https://doi.org/10.1016/j.cogpsych.2011.06 .002
- Hebb, D. O. (1961). Distinctive features of learning in the higher animal. In J. F. Delafresnaye (Ed.), *Brain mechanisms and learning: A sympo*sium (pp. 37–51). Blackwell Scientific.
- Heyes, C. (2018). Enquire within: Cultural evolution and cognitive science. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373(1743), Article 20170051. https://doi.org/10.1098/rstb.2017 .0051
- Humes, L. E., & Floyd, S. S. (2005). Measures of working memory, sequence learning, and speech recognition in the elderly. *Journal of Speech, Language, and Hearing Research*, 48(1), 224–235. https://doi .org/10.1044/1092-4388(2005/016)
- Hunt, R. H., & Aslin, R. N. (2001). Statistical learning in a serial reaction time task: Access to separable statistical cues by individual learners. *Journal of Experimental Psychology: General*, 130(4), 658–680. https:// doi.org/10.1037/0096-3445.130.4.658
- Hunt, R. H., & Aslin, R. N. (2010). Category induction via distributional analysis: Evidence from a serial reaction time task. *Journal of Memory* and Language, 62(2), 98–112. https://doi.org/10.1016/j.jml.2009.10.002
- Huron, D. (2006). Sweet anticipation: Music and the psychology of expectation. MIT Press. https://doi.org/10.7551/mitpress/6575.001.0001
- Isbilen, E. S., McCauley, S. M., Kidd, E., & Christiansen, M. H. (2020). Statistically induced chunking recall: A memory-based approach to statistical learning. *Cognitive Science*, 44(7), Article e12848. https://doi .org/10.1111/cogs.12848
- Itthipuripat, S., Cha, K., Byers, A., & Serences, J. T. (2017). Two different mechanisms support selective attention at different phases of training. *PLoS Biology*, 15(6), Article e2001724. https://doi.org/10.1371/journal .pbio.2001724
- Johansson, T. (2009). Strengthening the case for stimulus-specificity in artificial grammar learning: No evidence for abstract representations with extended exposure. *Experimental Psychology*, 56(3), 188–197. https:// doi.org/10.1027/1618-3169.56.3.188
- Karpicke, J. D., & Pisoni, D. B. (2004). Using immediate memory span to measure implicit learning. *Memory & Cognition*, 32(6), 956–964. https://doi.org/10.3758/bf03196873
- Kim, R., Seitz, A., Feenstra, H., & Shams, L. (2009). Testing assumptions of statistical learning: Is it long-term and implicit? *Neuroscience Letters*, 461(2), 145–149. https://doi.org/10.1016/j.neulet.2009.06.030
- 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
- Krishnan, S., Carey, D., Dick, F., & Pearce, M. T. (2020). Effects of statistical learning in passive and active contexts on reproduction and recognition of auditory sequences. osf.io/t7k34
- Krishnan, S., Leech, R., Aydelott, J., & Dick, F. (2013). School-age children's environmental object identification in natural auditory scenes: Effects of masking and contextual congruence. *Hearing Research*, 300, 46–55. https://doi.org/10.1016/j.heares.2013.03.003

- Lafond, D., Tremblay, S., & Parmentier, F. (2010). The ubiquitous nature of the Hebb repetition effect: Error learning mistaken for the absence of sequence learning. *Journal of Experimental Psychology: Learning, Memory,* and Cognition, 36(2), 515–522. https://doi.org/10.1037/a0018469
- Leech, R., Gygi, B., Aydelott, J., & Dick, F. (2009). Informational factors in identifying environmental sounds in natural auditory scenes. *The Journal of the Acoustical Society of America*, 126(6), 3147–3155. https://doi.org/10.1121/1.3238160
- Lim, S. J., Fiez, J. A., & Holt, L. L. (2019). Role of the striatum in incidental learning of sound categories. *Proceedings of the National Academy* of Sciences of the United States of America, 116, 4671–4680. https://doi .org/10.1073/pnas.1811992116
- Loui, P., Wessel, D. L., & Hudson Kam, C. L. (2010). Humans rapidly learn grammatical structure in a new musical scale. *Music Perception*, 27(5), 377–388. https://doi.org/10.1525/mp.2010.27.5.377
- Mareschal, D., & French, R. M. (2017). TRACX2: A connectionist autoencoder using graded chunks to model infant visual statistical learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1711), Article 20160057. https://doi.org/10.1098/rstb.2016.0057
- McCauley, S. M., & Christiansen, M. H. (2019). Language learning as language use: A cross-linguistic model of child language development. *Psychological Review*, 126(1), 1–51. https://doi.org/10.1037/rev0000126
- McDermott, J. H., Schemitsch, M., & Simoncelli, E. P. (2013). Summary statistics in auditory perception. *Nature Neuroscience*, 16(4), 493–498. https://doi.org/10.1038/nn.3347
- Miller, G. A. (1956). The magical number seven plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63(2), 81–97. https://doi.org/10.1037/h0043158
- Miller, G. A. (1958). Free recall of redundant strings of letters. Journal of Experimental Psychology, 56(6), 485–491. https://doi.org/10.1037/ h0044933
- Misyak, J. B., Christiansen, M. H., & Tomblin, J. B. (2010a). On-line individual differences in statistical learning predict language processing. *Frontiers in Psychology*, 1, Article 31. https://doi.org/10.3389/fpsyg .2010.00031
- Misyak, J. B., Christiansen, M. H., & Tomblin, J. B. (2010b). Sequential expectations: The role of prediction-based learning in language. *Topics* in Cognitive Science, 2(1), 138–153. https://doi.org/10.1111/j.1756 -8765.2009.01072.x
- Mitchel, A. D., & Weiss, D. J. (2011). Learning across senses: Crossmodal effects in multisensory statistical learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(5), 1081–1091. https://doi.org/10.1037/a0023700
- Newport, E. L., & Aslin, R. N. (2004). Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology*, 48(2), 127–162. https://doi.org/10.1016/S0010-0285(03)00128-2
- Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology*, 19(1), 1–32. https://doi.org/10.1016/0010-0285(87)90002-8
- Page, M. P. A., Cumming, N., Norris, D., Hitch, G. J., & McNeil, A. M. (2006). Repetition learning in the immediate serial recall of visual and auditory materials. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*(4), 716–733. https://doi.org/10.1037/0278 -7393.32.4.716
- Page, M. P. A., & Norris, D. (2009). A model linking immediate serial recall, the Hebb repetition effect and the learning of phonological word forms. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1536), 3737–3753. https://doi.org/10.1098/rstb.2009.0173
- Pearce, M. T. (2018). Statistical learning and probabilistic prediction in music cognition: Mechanisms of stylistic enculturation. Annals of the New York Academy of Sciences, 1423(1), 378–395. https://doi.org/10 .1111/nyas.13654
- 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
- Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, 36(4), 329–347. https://doi.org/10.1017/S0140525X12001495
- Pisoni, D. B., & Cleary, M. (2004). Learning, memory, and cognitive processes in deaf children following cochlear implantation. In F. G. Zeng, A. N. Popper, & R. R. Fay (Eds.), *Cochlear implants: Auditory prostheses and electric hearing* (Vol. 20, pp. 377–426). Springer. https://doi .org/10.1007/978-0-387-22585-2\_9
- Reber, A. S. (1967). Implicit learning of artificial grammars. Journal of Verbal Learning & Verbal Behavior, 6(6), 855–863. https://doi.org/10 .1016/S0022-5371(67)80149-X
- Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General*, 118(3), 219–235. https://doi.org/10.1037/0096-3445.118.3.219
- Redington, M., & Chater, N. (2002). Knowledge representation and transfer in artificial grammar learning. In R. M. French & A. Cleeremans (Eds.), *Implicit learning and consciousness: An empirical, philosophical and computational consensus in the making* (pp. 121–143). Psychology Press.
- Rey, A., Bogaerts, L., Tosatto, L., Bonafos, G., Franco, A., & Favre, B. (2020). Detection of regularities in a random environment. *The Quarterly Journal of Experimental Psychology*, 73(12), 2106–2118. https:// doi.org/10.1177/1747021820941356
- Robertson, E. M. (2007). The serial reaction time task: Implicit motor skill learning? *The Journal of Neuroscience*, 27(38), 10073–10075. https:// doi.org/10.1523/JNEUROSCI.2747-07.2007
- Rohrmeier, M., Rebuschat, P., & Cross, I. (2011). Incidental and online learning of melodic structure. *Consciousness and Cognition*, 20(2), 214–222. https://doi.org/10.1016/j.concog.2010.07.004
- Romberg, A. R., & Saffran, J. R. (2010). Statistical learning and language acquisition. WIREs Cognitive Science, 1(6), 906–914. https://doi.org/10 .1002/wcs.78
- Saffran, J. R. (2002). Constraints on statistical language learning. *Journal of Memory and Language*, 47(1), 172–196. https://doi.org/10.1006/jmla .2001.2839
- 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., & Griepentrog, G. J. (2001). Absolute pitch in infant auditory learning: Evidence for developmental reorganization. *Developmental Psychology*, 37(1), 74–85. https://doi.org/10.1037/0012-1649.37.1.74
- Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, 70(1), 27–52. https://doi.org/10.1016/S0010-0277(98)00075-4
- Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, 35(4), 606–621. https://doi.org/10.1006/jmla.1996.0032
- Saffran, J. R., Reeck, K., Niebuhr, A., & Wilson, D. (2005). Changing the tune: The structure of the input affects infants' use of absolute and relative pitch. *Developmental Science*, 8(1), 1–7. https://doi.org/10.1111/j .1467-7687.2005.00387.x
- Shanks, D. R., & Perruchet, P. (2002). Dissociation between priming and recognition in the expression of sequential knowledge. *Psychonomic Bulletin & Review*, 9(2), 362–367. https://doi.org/10.3758/bf03196294

- Shanks, D. R., & St. John, M. F. (1994). Characteristics of dissociable human learning systems. *Behavioral and Brain Sciences*, 17(3), 367–395. https://doi.org/10.1017/S0140525X00035032
- Siegelman, N., Bogaerts, L., Armstrong, B. C., & Frost, R. (2019). What exactly is learned in visual statistical learning? Insights from Bayesian modeling. *Cognition*, 192, Article 104002. https://doi.org/10.1016/j .cognition.2019.06.014
- Siegelman, N., Bogaerts, L., Christiansen, M. H., & Frost, R. (2017). Towards a theory of individual differences in statistical learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1711), Article 20160059. https://doi.org/10.1098/rstb.2016.0059
- Siegelman, N., Bogaerts, L., Elazar, A., Arciuli, J., & Frost, R. (2018). Linguistic entrenchment: Prior knowledge impacts statistical learning performance. *Cognition*, 177, 198–213. https://doi.org/10.1016/j.cognition .2018.04.011
- Siegelman, N., Bogaerts, L., Kronenfeld, O., & Frost, R. (2018). Redefining "learning" in statistical learning: What does an online measure reveal about the assimilation of visual regularities? *Cognitive Science*, 42(Suppl. 3), 692–727. https://doi.org/10.1111/cogs.12556
- Siegelman, N., & Frost, R. (2015). Statistical learning as an individual ability: Theoretical perspectives and empirical evidence. *Journal of Memory and Language*, 81, 105–120. https://doi.org/10.1016/j.jml.2015.02.001
- Takahasi, M., Yamada, H., & Okanoya, K. (2010). Statistical and prosodic cues for song segmentation learning by Bengalese finches (*Lonchura striata* var. *domestica*). *Ethology*, *116*(6), 481–489. https://doi.org/10 .1111/j.1439-0310.2010.01772.x
- 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, 372(1711), Article 20160056. https://doi.org/10.1098/rstb.2016.0056

- Tierney, A. T., Bergeson-Dana, T. R., & Pisoni, D. B. (2008). Effects of early musical experience on auditory sequence memory. *Empirical Mu*sicology Review, 3(4), 178–186. https://doi.org/10.18061/1811/35989
- Tillmann, B., & McAdams, S. (2004). Implicit learning of musical timbre sequences: Statistical regularities confronted with acoustical (dis)similarities. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(5), 1131–1142. https://doi.org/10.1037/0278-7393.30.5.1131
- 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
- Turner, R. E. (2010). Statistical models for natural sounds. University College London.
- Wells, J. B., Christiansen, M. H., Race, D. S., Acheson, D. J., & MacDonald, M. C. (2009). Experience and sentence processing: Statistical learning and relative clause comprehension. *Cognitive Psychology*, 58(2), 250–271. https://doi.org/10.1016/j.cogpsych.2008.08.002
- Williams, J. N. (2020). The neuroscience of implicit learning. Language Learning, 70(S2), 255–307. https://doi.org/10.1111/lang.12405
- Willingham, D. B., Wells, L. A., Farrell, J. M., & Stemwedel, M. E. (2000). Implicit motor sequence learning is represented in response locations. *Memory & Cognition*, 28(3), 366–375. https://doi.org/10.3758/ bf03198552

Received November 12, 2019 Revision received March 29, 2021

Accepted April 15, 2021