BRIEF REPORT

Implicit sequence learning using auditory cues leads to modality-specific representations

Y. Catherine Han[1] · Paul J. Reber[1]

Abstract

The implicit acquisition of statistical information from the environment is considered a fundamental type of human learning. Paradigms using visually cued sequences have been frequently used to study implicit learning. However, learning sequences of auditory cues is likely to be important in domains such as language or music. In three experiments, we established a novel auditorily cued implicit perceptual-motor sequence learning paradigm to compare to traditional visually cued sequence learning and identify whether this type of learning generalizes across cue modality. Participants exhibited reliable sequence-specific learning to auditory cues in all three experiments (Experiments 1-3), which was generally not influenced by explicit knowledge (Experiment 2). In Experiment 3, a large drop in knowledge expression in the novel cue modality was observed, suggesting that the acquired implicit sequence knowledge depended largely on sensory-specific representations. Overall, auditorily cued learning was similar to, though proceeded faster than, learning in comparable visually cued sequence learning paradigms. Similarity between learning from cues in different sensory modalities suggests that there may be a common process for the automatic extraction of sequential statistical structure. However, the lack of robust transfer sequence knowledge across modalities argues against a purely domain-general learning mechanism for all kinds of sequences. By expanding quantitative methodologies to characterize sequence learning in the auditory domain, these findings illustrate the possibility of bridging research in sequence and statistical learning domains to identify common mechanisms of complex cognitive skill and language learning.

Keywords Implicit learning and memory · Modality effects · Sequence learning · Statistical learning

Introduction

The implicit, automatic extraction of statistical structure from experience is a fundamental type of human learning (Reber, 2013). Several highly effective paradigms for characterizing and quantifying this learning process have used visually cued motor sequences with covertly embedded statistical structure, e.g., repeating sequences (Nissen & Bullemer, 1987; Robertson, 2007) or probabilistic manipulation of transition probabilities (Howard et al., 2004; Hunt & Aslin, 2001). Within implicit learning research, improved performance arising from the statistical structure of experience is considered important for practice in a learning process that leads to increasingly skilled behavior (Reber, 2013). However, in the closely related research area of statistical learning, sensitivity to structure in auditory sequences is often studied to gain insight into language processing (Saffran & Kirkham, 2018).

Here, we report a novel auditorily cued variant of the visual Serial Interception Sequence Learning (SISL) task (Sanchez et al., 2010), which is similar to the well-studied Serial Reaction Time (SRT) task (Nissen & Bullemer, 1987). With this new design, we aim to expand methodologies associated with implicit learning to be closer to statistical learning, evaluate whether learning is similar within the auditory domain, and test whether sequence-specific information can be transferred across modalities. Auditory sequence learning naturally applies to domains of language or music, which involve rapidly paced, temporally precise sequences of information. While memory systems approaches have not been regularly applied to learning and memory in these areas, commonalities between implicit and statistical learning and their relevance to language have been increasingly considered (Batterink et al., 2015, 2019; Christiansen, 2019; Conway, 2020; Conway & Pisoni, 2008; Perruchet & Pacton, 2006).

✉ Paul J. Reber
preber@northwestern.edu

[1] Northwestern University, 2029 Sheridan Road, Evanston, IL 60208, USA

Auditorily cued motor responses have been implemented in SRT (Buchner et al., 1997; Conde et al., 2012; Dennis et al., 2006; Goschke et al., 2001; Morin-Parent et al., 2017; Perruchet et al., 1997; Riedel & Burton, 2006; Zhuang et al., 1998) but not within SISL. The SISL task potentially provides a measure of implicit learning that is less contaminated by concomitant explicit memory in cognitively healthy participants. Most non-memory-impaired participants will recognize a covertly embedded repeating sequence after training (Willingham et al., 1993), which could theoretically drive performance via explicit anticipation. In contrast, SISL produces substantially less embedded sequence recognition across participants (Sanchez et al., 2010). Participants can even be provided full explicit knowledge before training without affecting implicit learning (Sanchez & Reber, 2013).

Using SISL, we can better address a key question about whether implicit sequence learning depends on a domain-general mechanism or results in knowledge representations tied to sensory modality. Our theoretical approach is grounded in the cognitive neuroscience of memory systems and the idea that operating characteristics of the learning process, such as transfer/flexibility, provide insight into neurocognitive bases of learning. In cognitively healthy (undergraduate) participant populations studied here, we hypothesize that intact learning of both implicit sequence knowledge and potentially some concomitant explicit knowledge may occur in parallel. Because SISL performance is fairly resistant to explicit knowledge influence (Sanchez & Reber, 2013; Experiment 2), the ability to apply sequence knowledge to a novel sensory modality should rely solely on implicit and not more flexible explicit memory. Within visually cued SISL paradigms, we previously used this approach to identify inflexible aspects of sequence learning (Sanchez et al., 2015).

Several candidates for a domain-general sequence learning mechanism have been proposed to depend on the basal ganglia (Seger, 2006), prefrontal cortex (Conway, 2020), or medial temporal lobe (Frost et al., 2015). However, some of these regions may reflect the operation of conscious, explicit, flexible learning mechanisms separate from implicit learning (Reber, 2013). If sequence knowledge can be applied across sensory modalities, this would argue for reliance on a purely domain-general sequential statistical learning process supported by one of these neural systems. A lack of transfer across modalities will indicate dependence on modality-specific knowledge in this implicit learning task, further ruling out the idea that SISL learning is purely motoric, consistent with previous SISL (Sanchez et al., 2015) and SRT reports (Dennis et al., 2006; Willingham, 1999; Willingham et al., 1989). The importance of sensory modality supports the broader prediction that plasticity dependent on sensory cortical areas plays an important role in sequential learning.

The set of studies reported here includes three experiments, the first of which establishes robust sequence-specific learning to auditory cues via SISL (Experiment 1). In Experiment 2, we tested whether providing explicit knowledge of the covertly embedded repeating sequence improved measures of implicit sequence knowledge. Experiment 3 tested whether participants could transfer their acquired sequence knowledge across visual and auditory modalities.

Experiment 1

Methods

Participants

A total of 34 participants were tested, including 28 participants recruited from the Northwestern Paid Participant Registry and six Northwestern undergraduate students enrolled in the introductory psychology course. Paid participants were compensated $15 for their participation, and students received course credit for their participation.

Of 34 participants, 26 (76%) completed the session, with five not completing the post-session recognition tasks due to a computer error. Seven participants could not complete the session within the 1.5-h protocol. One additional participant was excluded for task non-compliance, reflected by extreme over-responding (> 1000 responses per 180-trial sub-block in five sub-blocks, < 3% accurate responses). Analyses were conducted for the full sample, except for the recognition task analyses, which were conducted with the 21 participants who completed the recognition portion.

Sample Size Justification

In three recently published works with visually cued SISL, we observed sequence-specific learning effects (see Experimental Paradigm below) of a $SSPA = 10.10\%$, $SD = 9.66\%$, Cohen's $d = 1.04$ (Sanchez & Reber, 2013, implicit condition), $SSPA = 14.82\%$, $SD = 10.53\%$, Cohen's $d = 1.41$ (Thompson et al., 2014 Experiment 1 non-depletion condition), and $SSPA = 16.28\%$, $SD = 7.60\%$, Cohen's $d = 2.14$ (Sanchez et al., 2015, Experiment 3 standard condition). Assuming the learning effect would be similar with auditory cues, we estimated that we would have > 95% power to detect a reliable learning effect with a sample of 30 participants.

Materials

The task layout was presented on a computer monitor (23", 1920 x 1080 pixel resolution). The task was presented within a frame of 600 x 800 pixels (~ 60 cm typical viewing distance, ~ 20.0° visual angle for typical viewing distance). Sounds were presented binaurally via Steelseries Siberia P800 headphones, adjusted to a comfortable listening level for each participant.

Experimental Paradigm: Auditory SISL Task

Auditory cues were used to signal one of four motor keypress responses (keyboard keys D, F, J or K, marked visually on screen) based on the pitch of the cue (Fig.1). Each auditory cue (single trial) was initially presented as a set of three 100-ms tones of the same frequency with 600 ms between tones (1500 ms total cue duration). Participants were instructed to press the key that corresponded to its matching pitch during the onset of the third tone and each keypress reflected a single trial for performance measures. A response was considered correct if the key was pressed within 300 ms of the third tone. Feedback was provided visually by the target circle corresponding to the keypress, flashing green for correct responses and red for incorrect responses. The next cue started 1800 ms after the beginning of the previous cue (300 ms from cue offset to next cue onset).

As in typical implicit sequence learning paradigms, the auditory cues followed a covertly embedded, repeating, 12-item sequence (second-order conditional SOC structure; Reed & Johnson, 1994) during the majority of training trials (80%). Each 12-item sequence contained 36 tones (three tones per trial, 12 trials total). Each 60-trial sub-block contained four repetitions of the repeating SOC sequence (48 trials) and 12 trials of an unfamiliar SOC sequence, with the unfamiliar sequence occurring anywhere within the 60-trial sub-block. After training, test trial blocks contained repetitions of the same sequence (33% of test blocks) and two novel, unpracticed sequences (33% of test blocks for each novel sequence, 67% total). Trained and novel sequences were chosen randomly for each participant from the pool of 256 unique possible 12-item SOC sequences.

Cue duration and task timing were adaptively adjusted for each participant to target a performance level of 80% accuracy by speeding or slowing the task based on recent responses. After 12 responses, if 11 or 12 were correct, the duration of each auditory cue was shortened by ~ 5% (multiplied by 20/21), which was applied to the spacing between the three successive tones (which were always 100 ms). If nine or fewer responses were correct, the duration was increased by ~ 5% (multiplied by 21/20). This adaptive cue duration algorithm maintains a relatively constant level of non-ceiling performance to permit measurement of sequence learning as a Sequence Specific Performance Advantage (SSPA), defined as the average response accuracy to cues within the repeating sequence minus average accuracy for non-repeating sequence cues. As training progresses, SSPA generally increases, reflecting improved accuracy to the trained repeating sequence compared to non-repeating sequences. Adaptive cue duration adjustment results in similar overall task accuracy across all participants.

Task difficulty and individual differences in general task performance are measured by the cue duration at which performance stays at the 80% overall accuracy target.

Procedure

Participants were allotted 1.5 h to complete the protocol. Participants first completed a brief practice phase to learn the pitch to keypress mappings (Fig. 1). Participants listened to each of the four cues and practiced making the motor response to coincide with the third tone. Participants then completed three 540-trial training blocks (nine 60-trial sub-blocks, 36 12-item SOC sequence repetitions/block, 108 sequence repetitions total) and one 540-trial test block. Self-terminated breaks were provided after each 540-trial block, and no indication was provided to the participants as to whether they were performing a training or test block. To avoid underestimating the sequence-specific performance measure, the adaptive cue duration algorithm was disabled during the test block, maintaining the cue duration from the end of training. After completing training and test blocks, participants were informed that a repeating sequence had been present and then completed an explicit recognition test. Five different sequences composed of the same tones were presented (twice in succession), including the repeating sequence and four novel sequences. Each was rated on a scale from 1 to 9 (1 = not sure they heard the sequence during training, 5 = unsure, 9 = sure they heard the sequence during training). A recognition score for each participant was calculated as the difference between their rating for the practiced repeating sequence and their average rating for the four novel sequences.
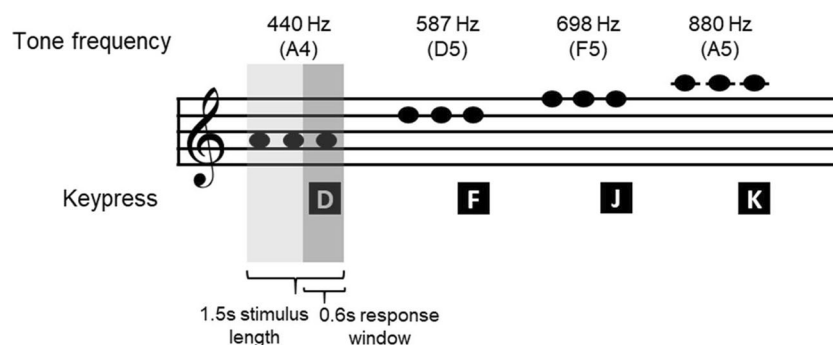

Fig. 1 Auditory SISL Task. Auditory cues were presented as three short tones at the indicated frequency associated with a specific keypress response (D, F, J, or K). Initial duration for a short tone sequence shown in light *grey*, response window shown in *dark grey*
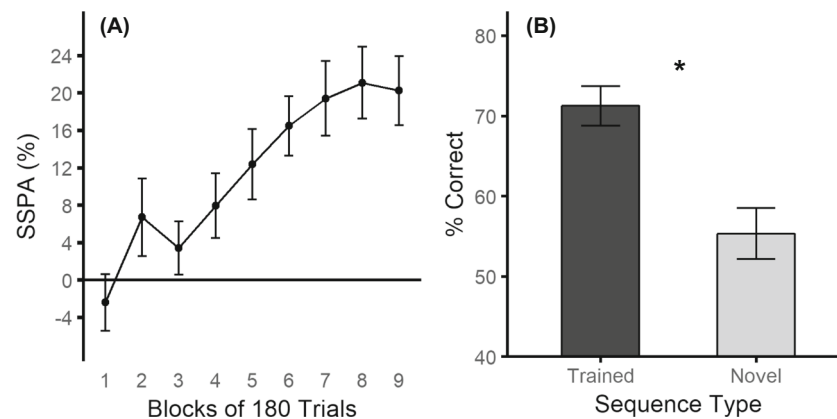
Fig. 2 **a** Increased repeating sequence knowledge emerged over training measured by higher accuracy for trials within the repeating sequence compared to non-repeating sequences (SSPA). **b** At test, participants were reliably more accurate in performance (% correct) during blocks of the trained repeating sequence compared to blocks of novel untrained sequences

Results

Sequence knowledge was compared across 180-trial training sub-blocks by a repeated measures ANOVA on SSPA, the difference in performance accuracy during repeating sequence trials (80%) and novel sequence trials (20%). SSPA increased across training in a significant linear trend, $F(1,25) = 44.26$, $p < .001$, $\eta^2 = .64$ (Fig. 2a). During test (Fig. 2b), participants exhibited higher accuracy for the trained repeating sequence ($M = 71.3\%$, $SE = 2.5\%$) compared to untrained novel sequences ($M = 55.4\%$, $SE = 3.2\%$), a SSPA of 15.9% ($SE = 3.1\%$) that was reliably greater than chance (0% SSPA), $t(25) = 5.22$, $p < .001$, 95% CI [9.64%, 22.18%], $d = 1.02$. In addition to SSPA, the average cue duration decreased across training blocks, $F(1,25) = 86.31$, $p < .001$, $\eta^2 = .78$, settling to an average cue duration of 0.9s/cue ($SE = 0.1$s) at test.

Participants gave higher recognition ratings for the practiced sequence ($M = 7.2$, $SE = 0.4$; 9-point scale) compared to novel ($M = 5.5$, $SE = 0.3$), mean difference ($M = 1.7$, $SE = 0.5$), $t(20) = 3.40$, $p < .01$, 95% CI [0.64, 2.68], $d = 0.74$. Recognition scores did not reliably correlate with test SSPA scores ($r = .25$, $p = .27$, $BF = 0.74$), nor was there a reliable difference in test SSPAs between participants with recognition scores above (SSPA: $M = 17.5\%$, $SE = 5.8\%$; Recognition: $M = 3.1$, $SE = 0.2$) and below the median (SSPA: $M = 17.0\%$, $SE = 4.7\%$; Recognition: $M = 0.06$, $SE = 0.7$), $t(19) = -0.07$, $p = .94$, 95% CI [– 16.21%, 15.13%], $d = 0.03$. The estimated Bayes factor for the correlation did not show evidence in favor of the alternate hypothesis ($BF = 0.74$). Specifically, the data were approximately 0.74 to 1 in favor of the alternate hypothesis over the null, which is usually considered weak evidence for observable differences between groups ($BF > 3$ for evidence against H_0) (Kass & Raftery, 1995). Cohen's d for below median recognition scores was 0.03, where 0.2 is traditionally considered a small effect size, 0.5 a moderate effect, and 0.8 or greater as a large effect (Cohen, 1992). Participants with below median recognition scores exhibited virtually no

recognition of the repeating sequence, $t(9) = 0.082$, $p = .93$, 95% CI [– 1.53, 1.65], $d = 0.03$, but produced a robust increase in accurate responding to cues within the repeating sequence, $t(9) = 3.61$, $p < .01$, 95% CI [6.34%, 27.56%], $d = 1.14$.

Experiment 1 Discussion

Participants exhibited robust sequence learning to auditory cues, paralleling prior visually cued studies. Precisely timed motor responses to the third tone in a three-tone set were reliably more accurate when the responses occurred within the covertly embedded repeating sequence. The level and rate of sequence learning observed here (16% SSPA) was somewhat larger than prior visually cued studies (Sanchez et al., 2010; Sanchez & Reber, 2012). In prior work, SSPA has generally been found to be linearly related to the logarithm of sequence repetition amount. By this formula, we would have expected a SSPA here of ~ 9% [7.5%, 10.5% CI] for the 108 sequence repetitions used in training. The speed at which the task was administered here was somewhat slower than visually cued SISL, leading to slightly higher exclusion of participants due to being unable to complete the task (and this exclusion rate was still lower than an attempt to use a different type of auditory cue[1]).

---

[1] *Experiment 1a.* A partial replication of Experiment 1 was completed with 23 participants who completed a similar auditory SISL protocol. A glissando-like effect at cue offset was used to signal the motor response timing, which was constructed by a rising pitch (20% of original frequency). Participants attempted to respond to the offset of the cue based on pitch contour, which proved difficult for participants with only 14 (61%) completing the protocol within the 1-h session. However, these participants exhibited reliable sequence knowledge via test block SSPA ($M = 13.0\%$, $SE = 5.0\%$, $t(13) = 2.60$, $p < .05$, $d = 0.70$), providing additional evidence of reliable sequence learning to a different auditory cue, despite the high exclusion rate raising concerns about reliability. We briefly note this result to raise the idea of other auditory methods to signal the motor response timing for future extensions of this research and as a cautionary note on the perceptual difficulty of some alternate approaches.

Experiment 2

While Experiment 1 demonstrated reliable auditorily cued sequence learning, participants exhibited some explicit knowledge of the sequence after practice. Despite patient work demonstrating memory dissociation of implicit and explicit memory, the separability of memory types in cognitively healthy participants has been debated for over fifty years (Reber, 2013). Common approaches necessarily depend on null findings, such as an absolute absence of explicit memory (methodologically intractable due to the need to prove a universal null; Merikle, 1994) or absence of correlations between recognition and implicit learning scores, as seen in Experiment 1. Memory systems theory does not require that either of these findings be null, as cognitively healthy participants will remember aspects of the task even if explicit memory did not support performance. To address this, we reported an improved methodology (Sanchez & Reber, 2013) in which participants were provided full explicit sequence knowledge prior to visually cued SISL learning. In SISL, the fast-paced nature of the task makes using explicit sequence knowledge difficult, and its availability did not lead to better sequence-specific performance. In Experiment 2, we used this approach with auditorily cued sequences to evaluate the influence of explicit memory on this new version of the task.

Methods

Participants

Participants were 73 Northwestern undergraduate students who received course credit (Introduction to Psychology) for their participation. Of 73 recruited participants, 51 (70%) completed the session through the test block (Explicit Instruction $n = 26$; Implicit $n = 25$), with one (Implicit) not completing the post-session recognition and recall tasks (see Procedure below). For the 22 excluded participants, 18 were unable to complete the session within the 1.5-h protocol ($n = 10$ Implicit, $n = 8$ Explicit). Three participants did not finish due to computer error. One participant left early due to a scheduling error.

Materials

The SISL task presentation was nearly identical to Experiment 1. Minor changes were made to the adaptive cue duration algorithm to allow participants to reach their performance speed ($\sim 80\%$ accuracy) more quickly. Performance was evaluated after every six trials instead of 12, where cue duration was shortened by $\sim 5\%$ (multiplied by 20/21) if 6 were correct, which applied to the spacing between the three successive tones (each tone was always 100 ms). If four or fewer responses were correct, the duration was increased by $\sim 5\%$

(multiplied by 21/20). In addition to changes in adaptive cue duration, the response window was shortened so that participants were required to make their responses within ~ 190 ms instead of ~ 300 ms.

Procedure

As in Experiment 1, participants first completed a brief practice to learn the pitch-to-keypress mappings. Participants were then randomly assigned to receive explicit instruction (Explicit, $n = 38$ tested, $n = 26$ included) or remain naïve to the repeating sequence (Implicit, $n = 35$ tested, $n = 25$ included). Participants in the Implicit condition did not receive any information regarding the embedded 12-item repeating sequence and performed the task following the same procedure as Experiment 1. Participants in the Explicit condition were told there was an embedded repeating sequence and were instructed to memorize their sequence prior to completing SISL training. Memorization was done by listening to the entire sequence and then repeating back (by keyboard responses) the entire sequence a total of five times. To ensure that participants in the Explicit condition maintained robust explicit sequence knowledge throughout the session, participants completed three additional trials of listening and recalling the repeating sequence between each of the four training blocks (12 trials total). Including the five pre-training sequence memorization trials, participants completed 17 memorization trials total.

Participants completed four 360-trial training blocks (six 60-trial sub-blocks, 24 12-item SOC sequence repetitions/block, 96 sequence repetitions total) and one 360-trial test block, each with the same internal structure as Experiment 1. Slightly shorter training and test blocks were used in Experiment 2 to ensure the protocol could be completed within the 1.5-h session. Self-terminated breaks were provided after each block of 360 trials.

As in Experiment 1, cue speed was adjusted adaptively to target a consistent 80% correct level for the general task. The adaptive cue duration algorithm was disabled during the test block to avoid underestimating the sequence-specific performance measure, but no other indication was provided to the participants that they were performing a test block.

After the test blocks, participants completed explicit recognition and recall tasks. The recognition task was assessed after the session in the same manner as Experiment 1, and participants provided their confidence ratings for the repeating sequence and four novel foil trials on a scale of 0–100 (0 = not sure they heard the sequence during training, 50 = unsure, 100 = sure they heard the sequence during training). The recognition memory score was calculated as a difference score between the confidence rating provided for the repeating sequence minus the average of confidence ratings for the novel sequences. For the recall task, participants were instructed to

provide their 12-item repeating sequence twice and type in up to 24 responses. The recall memory score was calculated by identifying the longest matching subsequence between the participants' response and the trained repeating sequence (Sanchez & Reber, 2013).

Results

Learning was assessed using a 2 (Explicit, Implicit) x 8 (training blocks) mixed ANOVA on SSPA (Fig. 3a). SSPA increased across training in a significant linear trend, $F(1,49) = 31.24, p < .001, \eta^2 = .39$. There was a reliable effect of training condition, $F(1,49) = 4.24, p = .04, \eta^2 = .08$, and a reliable interaction, reflecting the advantage for learning in the Explicit condition that appeared at the end of the training phase of the experiment, $F(1,49) = 7.35, p = .009, \eta^2 = .13$.

The effect of explicit knowledge on test performance was minimal (Fig. 3b), with participants exhibiting similar SSPAs in Explicit (SSPA = 20.6%, SE = 3.0%; trained M = 72.5%, SE = 3.2%; novel M = 51.9.0%, SE = 3.8%) and Implicit groups (SSPA = 19.6%, SE = 2.9%; trained M = 74.0%, SE = 2.2%; novel M = 54.4%, SE = 3.0%) that were both reliably greater than chance (0% SSPA), $ts > 6.8, ps < .001$. Test performance was not significantly different between conditions, indicating no effect of explicit knowledge on SSPAs, $t(49) = 0.25, p = 0.81$, 95% CI [– 7.36%, 9.41%], $d = 0.07, BF = 0.29$.

For post-test recognition (0–100 scale), participants recognized the memorized sequence more accurately in the Explicit (M = 49.5, SE = 7.5) compared to Implicit conditions (M = 22.4, SE = 7.4), $t(48) = 2.56, p < .05$, 95% CI [5.78, 48.31], $d = 0.72$ (Fig. 3c). Recognition scores (trained minus average of

untrained sequence confidence ratings) did not reliably correlate with test SSPA scores within Explicit ($r = .16, p = .44, BF = 0.54$) or Implicit conditions ($r = .024, p = .91, BF = 0.44$).

Post-test recall (amount of trained minus untrained sequence recalled) was above chance (difference score = 0) for both the Implicit ($M = 1.59, SE = 0.46, t(23) = 3.4, p < .01, d = 0.70$) and Explicit conditions ($M = 6.03, SE = 0.60, t(25) = 10.0, p < .001, d = 1.96$). There was a main effect of sequence type, condition, and an interaction ($F(1,49) > 22.0, ps < .001, \eta^2s > .31$), indicating that generated sequences matched trained rather than untrained sequences more for the Explicit than Implicit conditions. In the Explicit condition, recall training performance (amount of 12-item trained sequence recalled) increased from the five pre-training trials ($M = 5.92, SE = 0.26$) to the last recall training block ($M = 10.03, SE = 0.43$), meaning their memorization of the trained sequence improved over blocks. Additionally, 21 of 26 Explicit participants were capable of perfectly recalling their 12-item trained sequence during training. These participants, however, did not exhibit more sequence-specific knowledge on the test blocks than the Implicit group participants ($M = 24.4\%, SE = 3.0\%, t(45) = 1.16, p = 0.25$, 95% CI [– 3.5%, 13.1%], $d = 0.34, BF = 0.50$).

Task speed changes (measured as onset-to-onset ISI) were assessed by a 2 (condition) x 8 (training blocks) ANOVA, with average cue duration decreasing across training blocks, $F(1,49) = 180.02, p < .001, \eta^2 = .79$. Average cue duration settled to 0.9 s/cue (SE = 0.05 s) at test in the Explicit condition and 1.0 s/cue (SE = 0.08 s) in the Implicit condition. There was no significant effect of condition or interaction, $Fs < 0.83, ps > .36, \eta^2s < .01, BFs < 0.15$.
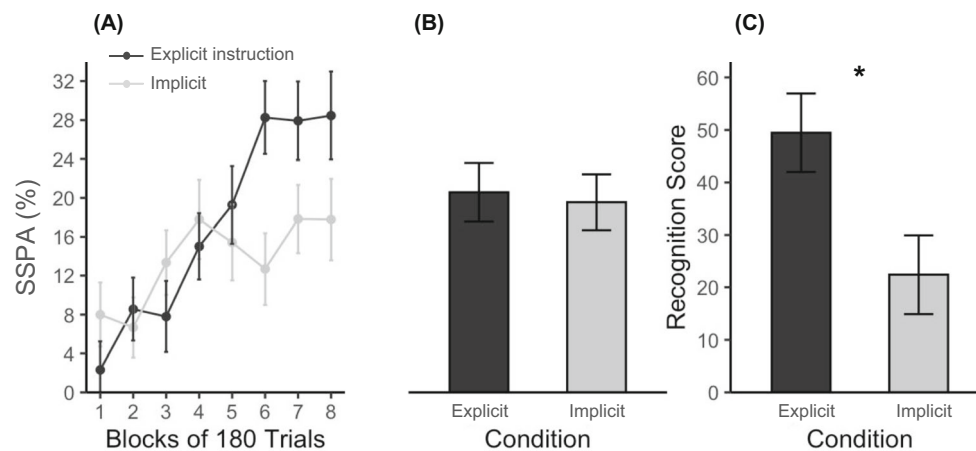

Fig. 3 a SSPA (Sequence-Specific Performance Advantage) indicates the difference score for the trained repeating sequence compared to novel sequence performance at test for the explicit training (*dark grey*) and implicit training conditions (*light grey*). Differential accuracy in performance for the trained repeating and novel sequences emerged over training for explicit and implicit training conditions. **b** Test performance

(SSPA in %) for the explicit training (*dark grey*) and implicit training (*light grey*) conditions were both reliably greater than chance (0% SSPA) but not significantly different from each other. **c** Explicit recognition scores for the sequence were reliably greater for those that received explicit instruction (*dark grey*) compared to those who did not (*light grey*). However, this did not correlate with sequence test performance (panel B)

Experiment 2 Discussion

Participants in both conditions exhibited robust learning of the embedded repeating sequence, replicating Experiment 1. As in our prior report (Sanchez & Reber, 2013), explicit instruction led to better explicit sequence knowledge on post-test recognition and recall measures but did not produce better sequence-specific performance at test. In contrast, we observed evidence for greater sequence-specific benefits at the end of training by Explicit participants (Fig. 3a), suggesting implicit and explicit knowledge potentially interact differently in auditorily cued SISL than in visual.

The similar performance on the test blocks for the groups with and without robust explicit knowledge of the repeating sequence is consistent with our prior findings that explicit knowledge does not lead to better sequence-specific performance on the SISL task. This is also consistent with the general lack of correlation between the degree of explicit knowledge and implicit scores (Experiment 1). Unlike these results, we observed better sequence-specific performance late in the training period for participants with full explicit knowledge of the sequence. It is unclear why this did not carry forward to the test block, which only differed in the percentage of trials that followed the repeating sequence. However, this finding suggests that auditory-cued sequence learning may lead to different interactions between implicit and explicit knowledge than visually cued SISL does.

The idea that auditory-based information supports more interaction between memory types than we observed with visually cued sequences suggests that these forms of sequence learning recruit domain-specific mechanisms rather than a single domain-general mechanism. In Experiment 3, we tested this hypothesis by examining transfer of sequence knowledge across sensory modalities. Transfer across sensory modalities would indicate the presence of a domain-general sequence learning mechanism. Lack of transfer implies inflexible, sensory-specific implicit learning.

Experiment 3

Methods

Participants

Participants were 70 Northwestern undergraduate students who received course credit (Introduction to Psychology) for their participation. Of 70 recruited participants, 56 (80%) completed the session (auditory $n = 29$, visual $n = 27$; including one participant who only completed two-thirds of the test). Thirteen participants were excluded for not completing the session within the 1.5-h protocol ($n = 12$ auditory, $n = 1$ visual). The remaining excluded participant (visual) did not complete the experiment due to computer error.

Materials

Auditory SISL condition This auditorily cued condition was identical to Experiment 2.

Visual SISL condition The visually cued condition was similar to that found in prior SISL studies (Sanchez et al., 2010). Task layout was on the same computer monitor within the laboratory as in Experiments 1 and 2. Participants observed circular cues (90 pixels in diameter, 2.4 cm) that appeared at the top of the display and moved vertically downwards towards one of four targets, 565 pixels away vertically, and spaced 200 pixels (5.3 cm) apart horizontally across the bottom of the screen. The starting cue velocity was 1.5 s/cue, such that the initial velocity was 300 pixels/second (7.9 cm/s) moving downwards from onset to the target on the screen. Cues were presented serially at the same rate as in the auditorily cued condition (1.8 s onset-to-onset ISI). Correct responses required a precisely timed correct keypress, coinciding with the cue moving through the bottom target zone. The response window was identical to that of the auditorily cued condition, 25% of the cue travel distance (190 ms) around the moment of exact cue-target overlap. After a response, visual feedback was provided (red/green flashes, as in the auditory conditions) and the cue was removed from the screen for correct responses for additional clarity about success to the participants.

Procedure

Participants were allotted 1.5 h to complete the protocol and were randomly assigned to complete training in the auditory or visual SISL conditions (Fig. 4). As in Experiments 1–2, participants completed a practice phase to learn the stimulus-to-keypress mappings immediately prior to auditory and visual versions of the task. They were not informed that they would be tested on both auditory and visual modalities. Participants then completed four 360-trial training blocks (96 12-item SOC sequence repetitions total) with self-terminated breaks between each block. This was followed by two 360-trial test blocks, one in the same sensory modality as training and the other in the untrained sensory modality (Fig. 4). Each test block contained the repeating, practiced sequence (120 trials) and two novel, unpracticed sequences (120 trials each). The presentation order of cue modality at test was counterbalanced. Given the time constraint and increased amount of test measures, a post-training explicit recognition task was not administered.

During test blocks in which the cue modality was different from training, cue duration was reset to 1.5 s/cue. Prior to the change in cue modality, participants completed a practice
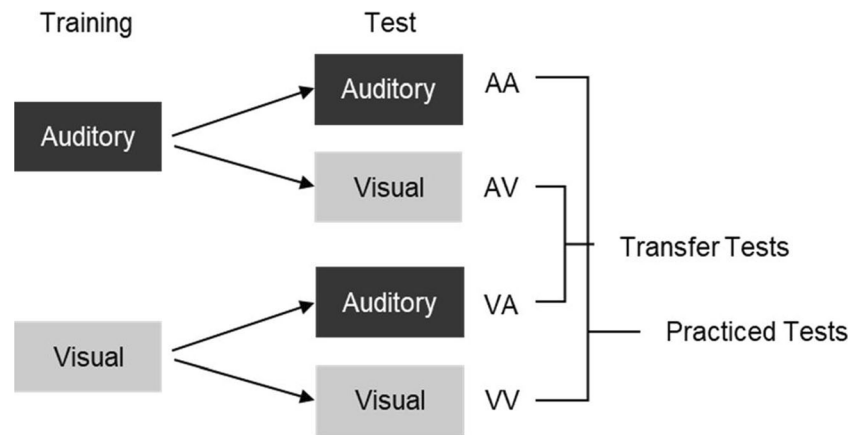
Fig. 4 Experiment 3 design. Participants completed training with either auditory or visual cues and completed tests with both auditory and visual cues. Practiced tests with the same modality as training were denoted as AA (auditory training, auditory test) and VV (visual training, visual test).

Transfer tests with a different sensory modality cue compared to training were denoted as AV (auditory training, visual test) and VA (visual training, auditory test)

phase to learn the stimulus-to-keypress mappings. Participants subsequently completed a 300-trial speed adjustment block (25 novel non-repeating SOC sequences) that re-attained the cue duration/speed for 80% accuracy in the new modality. This adjustment block was necessary to account for differences in difficulty between modalities without affecting sequence-specific performance. Subsequently, the adaptive cue duration algorithm was disabled during the test block to avoid underestimating the sequence-specific performance measure (SSPA).

Results

Learning was assessed using a 2 (training modality) x 8 (training blocks) mixed ANOVA on SSPA (Fig. 5a). SSPA increased in a significant linear trend, $F(1,54) = 27.44$, $p <$ .001, $\eta^2 = .34$. There was no reliable effect of training cue modality, $F(1,54) = 1.08$, $p = .30$, $\eta^2 = .020$, $BF = 0.72$, although there was a marginal interaction between training block and modality, reflecting a slight advantage for learning in the auditory cue condition, $F(1,54) = 3.54$, $p = .06$, $\eta^2 = .06$, $BF = 1.03$.

Transfer across modalities was assessed with a 2 x 2 x 2 mixed-model ANOVA comparing test modality (auditory/visual), training modality (auditory/visual), and presentation order (first/second) on test SSPA. Training modality, test modality, the interaction between training and test modality, and the interaction between test modality and order were all reliable, $Fs(1,52) > 7.25$, $ps <.01$, $\eta^2s > .12$. The training modality difference and interactions were related to the much larger SSPA observed for the AA compared to VV test conditions (collapsed across orders), $t(54) = 4.06$, $p < .001$, 95% CI [7.25%, 21.55%], $d = 1.06$ (Fig. 5b). All other effects were non-significant, $F(1,52) <$ 2.37, ps > .13, η^2s < .04, BFs < 0.82. Test order effects

were further examined by comparing transfer performance when the transfer condition (AV and VA) preceded the training modality condition (AA, VV). There was no difference in SSPA across orders, $t(54) = 1.78$, $p = .08$, 95% CI [– 0.53%, 9.68%], $d = 0.34$, $BF = 0.82$.

Reliable performance decrements were observed in the untrained sensory modality for both visual and auditory training, $F(1,54) > 9.37$, $ps < .01$, $\eta^2s > .15$. Post hoc t tests revealed that neither transfer condition produced SSPA reliably better than chance (0%), AV: SSPA = 3.3%, $SE = 1.9\%$ (trained $M =$ 74.5%, $SE = 2.3\%$; novel $M = 71.2\%$, $SE = 2.3\%$), $t(28) =$ 1.72, $p = .097$, 95% CI [– 0.64%, 7.26%], $d = 0.32$; VA: SSPA = 1.6%, $SE = 2.1\%$ (trained $M = 72.5\%$, $SE = 2.7\%$; novel $M =$ 70.9%, $SE = 2.2\%$), $t(26) = 0.74$, $p = .46$, 95% CI [– 2.74%, 5.84%], $d = 0.14$. The estimated Bayes factors indicated less evidence for transfer SSPAs being reliably greater than chance (AV: $BF = 0.72$; VA: $BF = 0.26$). The practiced conditions produced performances reliably greater than chance, AA: SSPA = 18.0%, $SE = 3.1\%$ (trained $M = 76.7\%$, $SE = 2.0\%$; novel $M = 58.7\%$, $SE = 3.2\%$), $t(28) = 5.82$, $p < .001$, 95% CI [11.66%, 24.33%], $d = 1.08$; VV: SSPA = 3.6%, $SE = 1.7\%$ (trained $M = 79.0\%$, $SE = 3.1\%$; novel $M = 75.4\%$, $SE =$ 2.6%), $t(26) = 1.83$, $p = .048$, 95% CI [0.03%, 7.16%], $d =$ 0.40. While there was less evidence for a robust learning signal for the VV condition ($BF = 1.3$) compared to AA ($BF >$ 150) based on Bayes factors, performance was generally worse in the untrained compared to trained modality.

Task speed changes (measured as onset-to-onset ISI) were assessed by a 2 (training modality) x 8 (training blocks) ANOVA. Modality, training block, and their interaction were all reliable ($Fs(1,54) > 29.81$, $ps < .001$, $\eta^2s > .36$), reflecting slower task speed for the auditory condition, speed increasing linearly across training blocks, and faster speedup for the visual modality. Participants were slower for the auditory condition and slower for the transfer auditory test (AA: $M = 0.9s$,
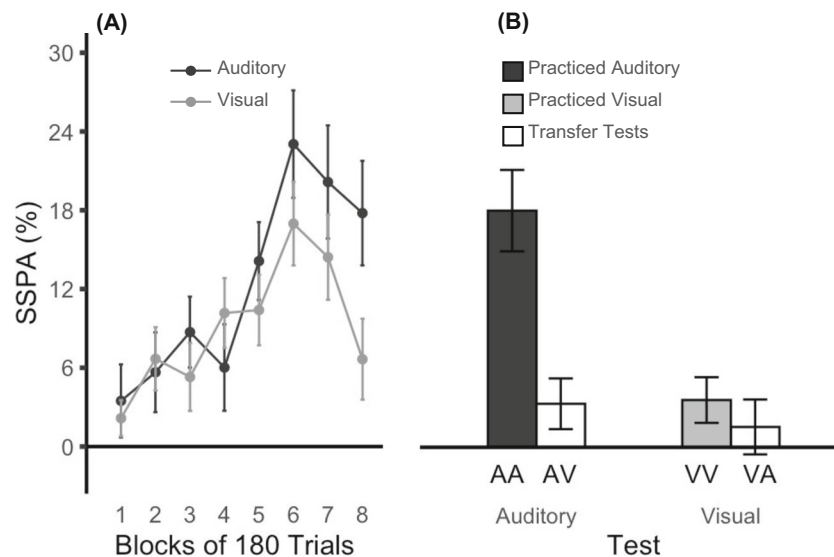
Fig. 5 a The SSPA (accuracy difference between trained and novel sequences) increased over training for both auditory cues (*dark grey*) and visual cues (*light grey*). **b** Participants completed implicit sequence knowledge tests with the practiced and novel cue modalities. Learning at test when the cues were in the same modality as training was reliable for both auditory cues (AA; *dark grey*) and visual cues (VV; *light grey*). A significant drop in learning occurred in the unpracticed modality regardless of the modality of the original learning (*bars in white*), indicating a relatively modality-specific representation of the sequence was acquired

SE = 0.1s; VA: M = 1.9s, SE = 0.2s), $t(54)$ = 4.92, p < .001, 95% CI [0.58 s, 1.41 s], d = 1.36. In contrast, the visual condition task speeds (VV: M = 0.6s, SE = 0.2s; AV: M = 0.5s, SE = 0.01s) were not reliably different, $t(54)$ = 1.02, p = .32, 95% CI [0.17s, 0.50s], d = 0.28.

Experiment 3 Discussion

Sequence learning was observed in both auditory and visually cued SISL, but the expression of knowledge was largely restricted to the modality in which it was acquired. A significant drop in sequence-specific performance was observed when participants performed the same sequence in the untrained modality. The lack of order effects indicates that this was not due to interference. Expression of sequence-specific knowledge was robust whether it preceded or followed the novel sensory modality. As observed in the prior experiments, SSPA associated with auditory cues was higher than prior reports with visual cues with a similar number of sequence repetitions, suggesting a general advantage for sequence learning within the auditory domain. The modality difference somewhat complicates the evaluation of transfer results, specifically in that the VA condition showed a more modest drop in SSPA, although reliable and quite substantial compared to learning in the AA and AV conditions. However, the inflexibility across sensory modalities was robust and statistically consistent across training modalities, indicating that sequence learning was largely sensory-specific despite the identical motor sequence being expressed.

General Discussion

Across three SISL experiments, we found reliable implicit learning of auditorily cued sequences. Participants became sensitive to repeating sequence structure, which improved their precisely timed motor response accuracy to a practiced order. In general, the sequence-specific advantage we measured in performance (SSPA) was higher for sequences cued auditorily than that found in similar paradigms using visual cues. This may reflect an advantage for the auditory domain in sequence learning, as suggested by Conway and Christiansen (2005). Experiment 3 provided evidence that SISL sequence learning reflects relatively modality-specific knowledge representations by showing that a large drop in sequence knowledge expression occurred in the untrained cue modality. This result suggests an important role for learning dependent on sensory regions and against a major role for a domain-general or purely motoric mechanism supporting all sequence learning. Results from all three experiments are consistent with our theory of a common learning process that extracts the statistical structure of behavior and operates independently in sensory-specific pathways (Conway, 2020; Reber, 2013).

In addition to larger implicit learning scores (SSPA), participants tended to score higher on recognition memory measures of the covertly embedded repeating sequence. Recognition memory did not correlate with higher implicit learning scores. When full explicit knowledge was provided, participants did not score higher on the final sequence learning test (Experiment 2). Unlike our prior research (Sanchez & Reber, 2013), participants exhibited some advantage during training with full explicit knowledge. While the increased

explicit memory effects were not substantial enough to raise the prospect that participants materially depended on explicit memory for performance, there may be a greater tendency for explicit sequence knowledge extraction within the auditory domain and more interactions with implicit learning during performance. This result, however, should be considered in the context of slower administration speeds for the auditory SISL task, which might have affected the availability of explicit knowledge (and led to somewhat higher exclusion rates based on inability to complete the task). If greater access to explicit knowledge is a feature of auditory-based sequence learning, it may reflect the fact that complex processes like language processing that are influenced by automatic statistical extraction still must produce conscious knowledge (comprehension). Considering the role of implicit learning in a complex process like language comprehension focuses attention on the need to better understand the interplay of implicit and explicit knowledge representations, an under-studied aspect of memory systems research.

The auditory SISL task here also provides a methodological connection between statistical and implicit learning research areas, which has been previously proposed (Batterink et al., 2015, 2019; Christiansen, 2019; Conway, 2020; Conway & Pisoni, 2008; Perruchet & Pacton, 2006). The reported importance of sensory modality in statistical learning (Conway & Christiansen, 2005, 2006; Frost et al., 2015) is consistent with our results and the idea of similar learning mechanisms supporting phenomena termed implicit or statistical learning. It is plausible that human auditory processing is particularly sensitive to sequential statistical structure, producing accelerated learning when auditory cues are used in both types of paradigms.

The idea that there is a common principle or underlying mechanism, however, must also be considered within the context that the paradigms are quite different. SISL requires rapid, precise, frequent motor responses to cues, which is not a feature of skills like language comprehension. While motor responses may influence learning process characteristics, Experiment 3 findings rule out the idea that SISL depended entirely on a motor (or domain-general) sequence learning representation, which would have supported robust transfer. The large drop in sequence-specific performance after the change in modality, even though the motor response sequence was identical, indicated that modality-specific representations were acquired during training. However, it should be noted that although sequence-specific performance was poor in transfer conditions, concluding there was no advantage at all requires sufficient power to accept a null hypothesis, which may not be available here. It is possible that there is a small contribution from a more domain-general process that will need to be examined in additional transfer studies.

While understanding the nature of implicit and explicit memory and their potential interaction in auditory sequence learning remain as areas of future research, the robust, auditorily cued, modality-specific, implicit learning found here supports the idea of a general plasticity mechanism that shapes representations and drives behavior. These findings suggest that implicit learning paradigms can be used to examine complex sequential learning in auditory domains, potentially serving as a bridge between implicit and statistical learning research areas towards a hypothesized general mechanism underlying extraction of complex statistical regularities. A unifying account of these research approaches will offer targeted guidance for future investigations of how statistical sequence knowledge is learned and stored in the brain, ultimately providing insight into how complex skills like language are learned.

References

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.

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.

Buchner, A., Steffens, M. C., & Rothkegel, E. E. & R. (1997). A multinomial model to assess fluency and recollection in a sequence learning task. *The Quarterly Journal of Experimental Psychology: Section A*, *50*(3), 631–663.

Christiansen, M. H. (2019). Implicit statistical learning: A tale of two literatures. *Topics in Cognitive Science*, *11*(3), 468–481.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159. https://doi.org/10.1037//0033-2909.112.1.155

Conde, V., Altenmüller, E., Villringer, A., & Ragert, P. (2012). Task-irrelevant auditory feedback facilitates motor performance in musicians. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00146

Conway, C. M. (2020). How does the brain learn environmental structure? Ten core principles for understanding the neurocognitive mechanisms of statistical learning. *Neuroscience & Biobehavioral Reviews*, *112*, 279–299. https://doi.org/10.1016/j.neubiorev.2020.01.032

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.

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.

Conway, C. M., & Pisoni, D. B. (2008). Neurocognitive basis of implicit learning of sequential structure and its relation to language processing. *Annals of the New York Academy of Sciences*, *1145*(1), 113–131.

Dennis, N. A., Howard, J. H., & Howard, D. V. (2006). Implicit sequence learning without motor sequencing in young and old adults. *Experimental Brain Research*, *175*(1), 153–164.

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.

Goschke, T., Friederici, A. D., Kotz, S. A., & Van Kampen, A. (2001). Procedural learning in Broca's aphasia: Dissociation between the implicit acquisition of spatio-motor and phoneme sequences. *Journal of Cognitive Neuroscience*, *13*(3), 370–388.

Howard, D. V., Howard, J. H., Japikse, K., DiYanni, C., Thompson, A., & Somberg, R. (2004). Implicit sequence learning: Effects of level of structure, adult age, and extended practice. *Psychology and Aging*, *19*(1), 79–92. https://doi.org/10.1037/0882-7974.19.1.79

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.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*(430), 773–795.

Merikle, P. M. (1994). On the futility of attempting to demonstrate null awareness. *Behavioral and Brain Sciences*, *17*(3), 412–412.

Morin-Parent, F., de Beaumont, L., Théoret, H., & Lepage, J.-F. (2017). Superior non-specific motor learning in the blind. *Scientific Reports*, *7*. https://doi.org/10.1038/s41598-017-04831-1

Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology*, *19*(1), 1–32.

Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: One phenomenon, two approaches. *Trends in Cognitive Sciences*, *10*(5), 233–238.

Perruchet, P., Bigand, E., & Benoit-Gonin, F. (1997). The emergence of explicit knowledge during the early phase of learning in sequential reaction time tasks. *Psychological Research*, *60*(1–2), 4–13.

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

Reed, J., & Johnson, P. (1994). Assessing implicit learning with indirect tests: Determining what is learned about sequence structure. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(3), 585.

Riedel, B., & Burton, A. M. (2006). Auditory sequence learning: Differential sensitivity to task relevant and task irrelevant sequences. *Psychological Research*, *70*(5), 337–344.

Robertson, E. M. (2007). The serial reaction time task: Implicit motor skill learning? *Journal of Neuroscience*, *27*(38), 10073–10075.

Saffran, J. R., & Kirkham, N. Z. (2018). Infant statistical learning. *Annual Review of Psychology*, *69*.

Sanchez, D. J., & Reber, P. J. (2012). Operating characteristics of the implicit learning system supporting serial interception sequence learning. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(2), 439.

Sanchez, D. J., & Reber, P. J. (2013). Explicit pre-training instruction does not improve implicit perceptual-motor sequence learning. *Cognition*, *126*(3), 341–351.

Sanchez, D. J., Gobel, E. W., & Reber, P. J. (2010). Performing the unexplainable: Implicit task performance reveals individually reliable sequence learning without explicit knowledge. *Psychonomic Bulletin & Review*, *17*(6), 790–796.

Sanchez, D. J., Yarnik, E. N., & Reber, P. J. (2015). Quantifying transfer after perceptual-motor sequence learning: How inflexible is implicit learning? *Psychological Research*, *79*(2), 327–343.

Seger, C. A. (2006). The basal ganglia in human learning. *The Neuroscientist*, *12*(4), 285–290.

Thompson, K. R., Sanchez, D. J., Wesley, A. H., & Reber, P. J. (2014). Ego depletion impairs implicit learning. *PLoS One*, *9*(10), e109370.

Willingham, D. B. (1999). Implicit motor sequence learning is not purely perceptual. *Memory & Cognition*, *27*(3), 561–572.

Willingham, D. B., Nissen, M. J., & Bullemer, P. (1989). On the development of procedural knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(6), 1047.

Willingham, D. B., Greeley, T., & Bardone, A. M. (1993). Dissociation in a serial response time task using a recognition measure: Comment on Perruchet and Amorim (1992). *Journal of Experimental Psychology: Learning, Memory, Cognition, 19*, pp. 1424–1430

Zhuang, P., Dang, N., Warzeri, A., Gerloff, C., Cohen, L. G., & Hallett, M. (1998). Implicit and explicit learning in an auditory serial reaction time task. *Acta Neurologica Scandinavica*, *97*(2), 131–137.