Perceiving structure in unstructured stimuli: Implicitly acquired prior knowledge impacts the processing of unpredictable transitional probabilities

Andrea Kóbor [a,+], Kata Horváth [b,c,d], Zsófia Kardos [a,e], Dezso Nemeth [c,d,f,*,+], Karolina Janacsek [c,d,*]

[a]: Brain Imaging Centre, Research Centre for Natural Sciences, Hungarian Academy of Sciences, *Magyar tudósok körútja 2, H–1117, Budapest, Hungary*

[b]: Doctoral School of Psychology, ELTE Eötvös Loránd University, *Izabella utca 46, H–1064, Budapest, Hungary*

[c]: Institute of Psychology, ELTE Eötvös Loránd University, *Izabella utca 46, H–1064, Budapest, Hungary*

[d]: Brain, Memory and Language Research Group, Institute of Cognitive Neuroscience and Psychology, Research Centre for Natural Sciences, Hungarian Academy of Sciences, *Magyar tudósok körútja 2, H–1117, Budapest, Hungary*

[e]: Department of Cognitive Science, Budapest University of Technology and Economics, *Egry József utca 1, H-1111, Budapest, Hungary*

[f]: Lyon Neuroscience Research Center (CRNL), INSERM, CNRS, Université de Lyon, *Centre Hospitalier Le Vinatier – Bâtiment 462 – Neurocampus 95 Boulevard Pinel 69675 Bron, Lyon, France*

*: These authors contributed equally to this work.

+: Corresponding authors:

Andrea Kóbor

Brain Imaging Centre, Research Centre for Natural Sciences, Hungarian Academy of Sciences; Magyar tudósok körútja 2., H–1117, Budapest, Hungary.

E-mail: kobor.andrea@ttk.mta.hu; Phone: + 36 1 382 66 15

Prof. Dr. Dezso Nemeth

CRNL – Lyon Neuroscience Research Center, Université Claude Bernard Lyon 1;

CRNL – Centre Hospitalier Le Vinatier – Bâtiment 462 – 95 Bd Pinel – 69675 Bron Cedex – Lyon, France.

E-mail: dezso.nemeth@univ-lyon1.fr; Phone: +33 4 81 10 65 46

Abstract

It is still unclear how implicit prior knowledge is involved in the extraction of the statistical structure underlying sensory input. Therefore, this study investigated whether the implicit knowledge on 2nd order transitional probabilities characterizing a stream of visual stimuli impacts the processing of unpredictable transitional probabilities embedded in a similar input stream. Young adults (N = 50) performed a four-choice reaction time (RT) task that consisted of structured and unstructured blocks. In the structured blocks, an alternating regularity repeated, yielding more probable and less probable short-range non-adjacent transitional probabilities. In the unstructured blocks, without the presence of the alternating regularity, the unique combinations of the short-range transitional probabilities occurred with equal probability, therefore, they were unpredictable. All task blocks were visually identical at the surface level. While one group of the participants completed first the structured blocks followed by the unstructured blocks, this was reversed in the other group. Participants completing first the structured blocks showed faster RTs to more probable than to less probable short-range transitional probabilities in both the structured and unstructured blocks, indicating the transfer of prior knowledge. In contrast, participants completing first the unstructured blocks showed this RT effect only in the structured blocks. These results overall suggest that, during the learning of serial dependence, the implicitly acquired prior knowledge on a transitional probability structure influences the processing of unstructured stimuli through the formation of internal models.

Keywords: implicit statistical learning, internal models, learning/memory transfer, prior knowledge, serial dependence, transitional probabilities

Introduction

Acquiring implicit knowledge on the statistical structure organizing environmental events is crucial for many cognitive functions and contributes to the automatization of behaviors (Armstrong, Frost, & Christiansen, 2017; Aslin, 2017; Kaufman et al., 2010; Maheu, Dehaene, & Meyniel, 2019). This ability involves not only the mere extraction of various statistical structures but also the efficient use of the acquired implicit knowledge across situations that differ in specific features at the surface level but share common features at the structural level. In everyday life, if conditions are substantially similar, we usually learn fast how to use the updated versions of applications or operating systems without checking manuals, running online searches, or even consciously accessing the course of our actions by transferring previous experiences. However, many aspects of how the generalization or transfer of the already acquired implicit knowledge takes place have remained unanswered (Robertson, 2018). Here, we focus on one aspect of this process in an implicit statistical learning situation: The transfer of the relationship between elements of a sequence (cf. Mosha & Robertson, 2016). Particularly, we investigate whether the implicit knowledge on transitional probabilities characterizing a stream of visual stimuli could be transferred across experimental task blocks that appear the same at the surface level but share common as well as distinct features at the structural level.

The processing of new information and the forming of expectations about future events are guided by those inferences that are based on prior experiences according to the broad frameworks of cognitive processing, learning, and decision making (e.g., Daw, Gershman, Seymour, Dayan, & Dolan, 2011; Friston, 2005; Friston, 2010; Friston, Stephan, Montague, & Dolan, 2014; Griffiths, Kemp, & Tenenbaum, 2008; Shohamy & Daw, 2015). This also pertains to implicit statistical learning. For instance, the experienced features of statistical structures spontaneously occurring in *random* sequences influenced randomness

perception and random sequence generation (Hahn & Warren, 2009; Sun et al., 2015; Sun & Wang, 2010; Warren, Gostoli, Farmer, El-Deredy, & Hahn, 2018). In addition, it was shown that the primarily experienced statistical structure limited the capacity to acquire the successive statistical structure (Gebhart, Aslin, & Newport, 2009). The persistency of the primarily acquired statistical structure and its influence on further processing were shown by event-related brain potentials, as well (Honbolygó & Csépe, 2013; Mullens et al., 2014; Todd, Provost, & Cooper, 2011). Findings of binary choice experiments also underpinned the relevance of implicit prior expectations. Namely, the so-called probability-matching behavior (i.e., matching choice probabilities to outcome probabilities instead of the optimal maximizing strategy) was linked to the tendency to search for an underlying statistical structure (e.g., Feher da Silva & Baldo, 2012; Gaissmaier & Schooler, 2008). It was also observed that the generation of sequence-wide expectations increased the frequency of this behavior (James & Koehler, 2011). Overall, previous findings highlight the role of prior internal models in the implicit processing of statistical structures. Meanwhile, the investigation of how the *direct* manipulation of these models influence the transfer of implicit knowledge across similar learning situations is still lacking.

Different statistical structures can be acquired from the same stimulus sequence (Conway & Christiansen, 2001; Daltrozzo & Conway, 2014). According to the model proposed by Meyniel, Maheu, and Dehaene (2016), instead of simpler statistics such as frequencies and alternations of events, the computation of time-varying, non-stationary, local transitional probabilities between consecutive events could be considered as the “building block” of implicit statistical learning and knowledge (see also Orbán, Fiser, Aslin, & Lengyel, 2008). This model explains a wide range of experimental results, including sequential effects in reaction time studies, the biased perception/generation of random sequences, novelty/surprise effects in EEG and fMRI studies, and the processes related to word

segmentation. In addition, accumulating evidence also suggests that sensitivity to simpler statistics and to more complex transitional probabilities are related to multiple time scales at the level of brain signals (Kóbor et al., 2018; Maheu et al., 2019). Altogether, these findings warrant the closer examination of the acquisition and transfer of transitional probabilities. Among the different transitional probabilities, humans have been found to be highly proficient in extracting the *non-adjacent* ones referring to predictive relations between elements of a sequence that includes ordered stimuli interspersed with random ones (Frost & Monaghan, 2016; Malassis, Rey, & Fagot, 2018; Mueller, Milne, & Männel, 2018; Rey, Minier, Malassis, Bogaerts, & Fagot, 2018). Therefore, here, we focus on the implicit processing of a sequence composed of 2nd order non-adjacent transitional probabilities, where elements in position n – 2 predict elements in position n.

From a methodological point of view, it is not obvious how one investigates what has been learned about the statistical structure underlying a given sequence or whether generalization has occurred. The study of Reed and Johnson (1994) suggests that to appropriately test whether the complex statistical structure per se has been learned, instead of a random transfer sequence, one should use training and transfer sequences that differ only in 2nd order transitional probabilities but are identical in terms of 1st order transitional probabilities and other simpler statistics (e.g., location frequency, transition frequency, reversal frequency, coverage, transition usage). By controlling for the latter characteristics of the training and transfer sequences, it can be ensured that the RT disruption across the sequences is due to acquiring the 2nd order transitional probability structure that changed from the training to the transfer sequence. The study of Reed and Johnson (1994), however, applied deterministic sequences, in which a certain stimulus was fully determined by the previous two stimuli. This leaves the question open how transfer occurs when more and less probable short-range relations can be identified as part of *probabilistic* sequences. Particularly, the

investigation of probabilistic sequences could provide insight into how implicit statistical learning operates in more general fields such as language acquisition and language processing (e.g., Christiansen, 2018; Saffran, Aslin, & Newport, 1996; Siegelman, Bogaerts, Christiansen, & Frost, 2017).

For this purpose, to measure the implicit acquisition and transfer of a 2nd order non-adjacent transitional probability structure, the present study used a specific four-choice reaction time (RT) task. This task, unknown to participants, in half of the task blocks, included an alternating regularity between non-adjacent trials yielding more probable and less probable short-range transitional probabilities (see Fig. 1). The short-range transitional probabilities were three successive trials, hereafter referred to as *triplets*. In the other half of the task blocks, the alternating regularity was absent but the unique triplets remained. The task blocks were labeled as *structured* and *unstructured* blocks, according to the presence or absence of the alternating regularity. While one group of the participants completed first the structured blocks followed by the unstructured blocks, the other group completed first the unstructured blocks followed by the structured blocks. By controlling for the probability of the triplets in the unstructured blocks, i.e., each appearing with equal probability, we tested whether acquisition has occurred at the level of transitional probabilities. If acquisition occurred this way, the RT difference between the more probable and less probable triplets would be comparable across the structured and unstructured blocks for participants completing the *structured blocks first*, because the primarily acquired implicit knowledge would have influenced the processing of stimuli in the further blocks. On the contrary, participants completing the *unstructured blocks first* could not become tuned to perceive the relations of stimuli primarily as triplets. Accordingly, their acquisition of the transitional probabilities would occur only over the structured blocks in a gradual manner.

Material and methods

Participants

Fifty healthy young adults took part in the experiment. They were undergraduate students from Budapest, Hungary. Participants had normal or corrected-to-normal vision, and according to the pre-defined inclusion criteria, none of them reported a history of any neurological and/or psychiatric condition, and none of them was taking any psychoactive medication. Half of the participants was randomly assigned to the Structured-first group (*n* = 25), while the other half was assigned to the Unstructured-first group (*n* = 25). The groups were differentiated by which half of the experimental task they started with; this is explained in the Procedure section below. Descriptive characteristics of participants in the two groups and their performance on standard neuropsychological tests are presented in Table 1. All participants provided written informed consent before enrollment and received payment (ca. 12 Euros) or course credit for taking part in the experiment. The study was approved by the United Ethical Review Committee for Research in Psychology (EPKEB) in Hungary and was conducted in accordance with the Declaration of Helsinki.

<PLEASE INSERT TABLE 1 HERE>

Experimental task

The Alternating Serial Reaction Time (ASRT) task. Implicit acquisition of 2nd order transitional probabilities was measured by a modified version of the ASRT task (Nemeth et al., 2010; Takács et al., 2018), which was optimized for a future fMRI study using a block design. In this task, a stimulus (a dog's head) appeared in one of four horizontally arranged empty circles on the screen (see Fig. 1C). Participants were instructed to press as quickly and accurately as possible one of the four response keys (Q, Y, M, or, O on a

QWERTZ keyboard) that corresponds to the position of the stimulus (Q = leftmost position [left index finger], Y = second position from left to right [left thumb], M = third position from left to right [right thumb], O = rightmost position [right index finger]). In this task version, participants were clearly informed about the unusual mapping between spatial positions and response keys in the task instruction. During a practice phase with at least two mini-blocks of fifteen random trials each, participants had the chance to practice these stimulus-response mappings until they felt confident in proceeding to the main task. (The experimenters also required them to achieve 98% of accuracy at least in the final mini-block).

In the original ASRT task, unbeknownst to participants, the stimuli are presented according to an eight-element sequence, within which predetermined/pattern (P) and random (r) elements alternate with one another. For instance, 2 – r – 1 – r – 3 – r – 4 – r is one of the sequences, where numbers denote the four predetermined positions on the screen from left to right and *r*s denote the randomly chosen positions out of the four possible ones (see Fig. 1A). There are 24 permutations of the four positions that could determine the applied sequence; however, because of the continuous presentation of the stimuli, there are only *six unique permutations*: 1 – r – 2 – r – 3 – r – 4 – r, 1 – r – 2 – r – 4 – r – 3 – r, 1 – r – 3 – r – 2 – r – 4 – r, 1 – r – 3 – r – 4 – r – 2 – r, 1 – r – 4 – r – 2 – r – 3 – r, and 1 – r – 4 – r – 3 – r – 2 – r. Note that each of these six permutations can start at any position; e.g., 1 – r – 3 – r – 4 – r – 2 – r and 2 – r – 1 – r – 3 – r – 4 – r are identical sequence permutations.

The alternating regularity yields a probability structure in which some chunks of three successive trials (triplets) occur more frequently than others. In the case of the 2 – r – 1 – r – 3 – r – 4 – r sequence, 2 – X – 1, 1 – X – 3, 3 – X – 4, and 4 – X – 2 triplets (X denotes the middle trial of the triplet) occur frequently since these triplets could have P – r – P or r – P – r structure. Meanwhile, for instance, 1 – X – 2 and 4 – X – 3 triplets occur less frequently since they could only have a r – P – r structure (see Fig. 1A). The former triplets are referred to as

high-probability triplets while the latter ones are referred to as *low-probability* triplets (e.g., Nemeth & Janacsek, 2011; Nemeth, Janacsek, Polner, & Kovacs, 2013). Construction of triplets could be considered as a method for identifying a hidden probability structure of the ASRT task. Namely, the final trial of a high-probability triplet is a probable (predictable) continuation for the first trial of the triplet while the final trial of a low-probability triplet is a less probable continuation. For instance, in the case of the above-mentioned sequence, if the first trial of a triplet is position 3, it is more likely (with 62.5% probability) to be followed by position 4 as the third trial than either position 1, 2, or 3 (with 12.5% probability each). Each trial (stimulus) is categorized as either the third trial of a high- or a low-probability triplet. Accordingly, the construction of triplets is applied as a moving window throughout the entire stimulus set: The third trial of a triplet is also the second trial of the following triplet, and so on; thus, all stimuli are categorized this way (Kóbor, Janacsek, Takács, & Nemeth, 2017; Kóbor et al., 2018; Szegedi-Hallgató et al., 2017). There are 64 possible triplets in the task: 16 of them are high-probability triplets and 48 are low-probability ones. With respect to the *unique* triplets, the third trials of high-probability triplets are five times more predictable based on the first trials than those of the low-probability triplets.

Generation and selection of the unstructured sequences. Besides the original *structured* ASRT sequences that included the alternating regularity, *unstructured* sequences, in which the alternating regularity was absent, were generated and selected to match the structured ASRT sequences in two aspects. *First*, both sequence types had to contain the same number of trials as they determined stimulus presentation in equal number of blocks. This meant the presentation of altogether 1920 triplets distributed over 48 blocks with 40 triplets in each, respectively (see below). *Second*, unstructured sequences had to contain the same 64 triplets as the structured sequences; however, the probability of occurrence of each unique triplet type had to be equal in the unstructured sequences. Therefore, each of the 64 triplets

had to occur 30 times (1920 triplets in total) in any of the generated unstructured sequences but without the presence and repetition of the alternating regularity.

For this purpose, *without* the use of the alternating regularity, several trial sets were generated in MATLAB 2015a (The MathWorks Inc., Natick, 224 MA, USA). Particularly, by randomly changing one trial of the trial sets at a time, the trial set minimizing the deviation from the optimal 30 times of occurrence was selected (the maximal error was set to two). Using this algorithm, a dozen of trial sets satisfying this criterion were kept. These trial sets were then subjected to three further constraints: (1) the maximal repetition of a unique triplet in any of the blocks could be no more than four; (2) the maximal immediate repetition of a trial [position] could be no more than five across the entire trial set; (3) in larger time bins (16 blocks) of the unstructured trial set, the overall occurrence probability of triplets that can be categorized as high- vs. low-probability in the structured ASRT sequences, should approximate 25% and 75%, respectively, since there are 16 unique high-probability and 48 unique low-probability triplets for a given ASRT sequence (see above the ASRT task description). This third constraint ensured that at the level of unique triplets, transitional probabilities were equal. Six of the trial sets were appropriate regarding constraints (1) and (2). Stimuli of these six trial sets were categorized into *triplets* following either of the six unique structured ASRT sequences (see Fig. 1B); and, constraint (3), i.e., the ratio of high- and low-probability triplets, was checked on these categorized trial sets. Finally, altogether 19 trial sets satisfied all three constraints and were kept using as *unstructured sequences*.

When assigning the structured ASRT and unstructured sequences to participants, we ensured that the distribution of the six unique ASRT sequence types was even across the two groups. The 1 – r – 2 – r – 3 – r – 4 – r sequence was used five times, all the other sequences were used four times in each of the groups (i.e., for 25 participants per group). For each participant, the selection of a sequence from the six unique types was pseudorandom. The

applied files containing the structured ASRT and unstructured sequences were matched one-to-one across the two groups. Note that for each respective participant, at the level of unique triplets, the identified high- and low-probability triplets were the same in the unstructured sequences as in the structured ASRT sequences (see Fig. 1A–B).

As a result of the procedure used for generating, selecting, and matching the sequences, in the present sample, the distribution of high- and low-probability triplets did not differ across the four stimulus positions either in the structured ASRT ($\chi^2(3) = 4.86$, $p = .183$) or in the unstructured sequences ($\chi^2(3) = 0.02$, $p = .999$); in addition, these associations between triplet distribution and stimulus position did not differ across the sequence types (Wald $\chi^2(3) = 2.34$, $p = .504$). When the high- and low-probability triplet categories were collapsed, the distribution of stimulus positions across sequence types also did not differ ($\chi^2(3) = 1.56$, $p = .670$). This way, lower-level characteristics of the sequences would not account for the assumed between-sequence RT variations related to acquiring the 2nd order transitional probability structure (cf. Reed & Johnson, 1994).

<PLEASE INSERT FIGURE 1 HERE>

Procedure

An experimental trial started with the presentation of the stimulus at one of the four positions for 500 ms. After stimulus offset, the image of the four positions was displayed for 200 ms. Then, the next trial started, yielding a 700-ms-long inter-trial-interval. The behavioral response (key-press) was expected during the whole trial from stimulus onset until the end of the trial (i.e., for altogether 700 ms, see Fig. 1C). These trial events were always the same with fix durations, irrespective of whether participants provided correct, incorrect, or missing response(s). In this task version, no feedback was presented as a function of the quality of the

response. The lack of feedback presentation and the fact that participants could proceed with the trial without providing the correct response ensured that each trial and each block had the same lengths, respectively. Importantly, only correctly responded trials were analyzed in the present study.

One block of the task contained 42 trials. There were 48 blocks with the structured ASRT sequence and 48 blocks with the unstructured sequence. In each of the structured blocks, the eight-element-long alternating regularity repeated five times after two starter trials that were not categorized as triplet elements (since also the foremost triplet technically required three successive trials). The alternating regularity was missing from the unstructured blocks, but, as in the structured blocks, 40 triplets followed the two starter trials that were not categorized as triplet elements. After each block, participants received feedback (lasting for 10, 12, or 14 secs [mean = 12 sec]) about their mean reaction time and accuracy in the given block. Altogether 96 blocks were completed (4032 trials in total).

The Structured-first group completed 48 structured blocks followed by 48 unstructured blocks. The Unstructured-first group completed 48 unstructured blocks followed by 48 structured blocks. All participants proceeded with the task from its structured/unstructured to unstructured/structured half without receiving information about any change in the task (see Fig. 1D). Two breaks (1.5 mins each) were inserted after the 32nd and 64th blocks, in which participants could have had a short rest. The experimental procedure lasted about 1.5 hours including the administration of a short post-task questionnaire. This assessed participants' task-solving strategies and their consciously accessible knowledge about the structure of the task and the transitional probabilities (Kóbor et al., 2017; Nemeth, Janacsek, & Fiser, 2013; Song, Howard, & Howard, 2007). Namely, participants were asked whether (1) they followed any task-solving strategies to improve their performance and (2) if yes, to what extent they found it efficient; (3) whether they noticed anything special regarding the task; (4) whether

they noticed any regularity in the sequence of stimuli; and (5) whether they noticed any substantial change in the sequence of stimuli. The first author (AK) qualitatively rated participants' answers to question (1) and (2), and rated the answers to questions (3), (4), and (5) on a 5-item scale (1 = "Nothing noticed", 5 = "Total awareness"). None of the participants reliably reported noticing the alternating regularity, the presence and repetitions of the triplets, or any change in the stimulus sequence between the task halves (the mean score for the three questions is 1.006, *SD* = 0.082). Although participants reported several strategies they found somewhat facilitating (e.g., counting the stimuli, fixating to the center of the screen, catching the rhythm of trials by silently singing, bouncing their legs, or moving their fingers), these were unrelated to the hidden structure of the task. Only one participant reported trying to search for some "logic" in the sequence but as a subjectively inefficient strategy.

The current ASRT task variant was written in and controlled by the MATLAB 2015a using the Psychophysics Toolbox Version 3 (PTB-3) extensions (Brainard, 1997; Pelli, 1997). Stimuli were displayed on a 15" LCD screen at a viewing distance of 100 cm. Neuropsychological tests (see section Participants) were administered a few days before the main experiment during a one-hour-long session.

Statistical analysis

Behavioral evidence accumulated so far indicates that participants respond increasingly faster to high-probability triplets compared with low-probability ones as the ASRT task progresses (e.g., Janacsek, Ambrus, Paulus, Antal, & Nemeth, 2015; Kóbor et al., 2017; Nemeth et al., 2010; Nemeth, Janacsek, Polner, et al., 2013; Takács et al., 2017; Tóth et al., 2017). Therefore, we measured whether *triplet learning* contrasting high- and low-probability triplets occurred over the course of the task on both the structured ASRT sequences and the unstructured sequences. Independent of triplet learning, we also measured *general skill*

improvements (faster RTs), which reflect more efficient visuomotor and motor-motor coordination due to practice (Hallgató, Győri-Dani, Pekár, Janacsek, & Nemeth, 2013).

Following the standard data analysis protocol established in previous studies using the ASRT task (e.g., J. H. Howard, Jr. & Howard, 1997; Kóbor et al., 2017; Nemeth, Janacsek, Polner, et al., 2013; Song et al., 2007; Virag et al., 2015), two types of low-probability triplets – repetitions (e.g., 1 – 1 – 1, 4 – 4 – 4) and trills (e.g., 1 – 2 – 1, 2 – 4 – 2, see Fig. 1B) – were eliminated from the analyses, because pre-existing response tendencies have often been shown to them (D. V. Howard et al., 2004). In addition, eight-block-long units of the task were collapsed into larger time bins labeled as *epochs*, yielding altogether six structured epochs (containing the ASRT sequence) and six unstructured epochs (containing the unstructured sequence). From this point of view, while the Structured-first group performed six structured epochs followed by six unstructured epochs, the Unstructured-first group performed six unstructured epochs followed by six structured epochs. Epochs are labeled consecutively in this paper (1, 2, etc.) within each sequence type. For each participant and epoch, separately for high- and low-probability triplets, median RT was calculated only for correct responses. Triplet learning *scores* in the structured as well as in the unstructured epochs were then calculated as the RT difference between triplet types (RTs to low-probability triplets minus RTs to high-probability triplets). *Overall* triplet learning *scores* were considered for the structured and unstructured epochs, respectively, as the mean of scores calculated for each of the six epochs.

Triplet learning was first quantified with a four-way mixed design analysis of variance (ANOVA) with Sequence (structured vs. unstructured), Triplet (high- vs. low-probability), and Epoch (1–6) as within-subjects factors and Group (Structured-first group vs. Unstructured-first group) as a between-subjects factor on the RTs. Second, to follow up the obtained effects and to more directly test the change in triplet learning as a function of the

different epoch/sequence types, three-way mixed ANOVAs with Triplet and Epoch as within-subjects factors and Group as a between-subjects factor were performed on the RTs related separately to the structured and unstructured epochs. Third, to test the change in triplet learning as a function of group assignment, three-way repeated measures ANOVAs with Sequence, Triplet, and Epoch as within-subjects factors were performed separately in the Structured-first and Unstructured-first group. Fourth, to follow up the previous group-wise analyses, ANOVAs with Triplet and Epoch as within-subjects factors were performed in the Structured-first group on the RTs related to the structured and the unstructured epochs, respectively; and the same two ANOVAs were conducted in the Unstructured-first group. In all ANOVAs, the Greenhouse-Geisser epsilon (ε) correction (Greenhouse & Geisser, 1959) was used when necessary. Original *df* values and corrected (if applicable) *p* values are reported together with partial eta-squared (η_p^2) as the measure of effect size. LSD (Least Significant Difference) tests for pair-wise comparisons were used to control for Type I error.

Regarding the possible experimental effects and their interpretation, it should be noted that the Triplet main effect implies triplet learning (faster RTs on high- than on low-probability triplets), the Triplet * Epoch interaction implies changes in triplet learning as the task progresses (usually an increase across epochs), and the Epoch main effect implies general skill (RT) improvements due to practice with the task. The *transfer effect* is indicated by the Sequence * Triplet * Group and/or the Sequence * Triplet * Epoch * Group interactions, as, if learning transfer occurred, triplet learning per se or its change over time should differ between the structured and the unstructured epochs *and* across the two groups experiencing the epoch/sequence types in a different order. In the Results section below, we use these terms when describing the observed statistical effects.

Results

Overall analysis

The Sequence (structured vs. unstructured) by Triplet (high- vs. low-probability) by Epoch (1–6) by Group (Structured-first group vs. Unstructured-first group) *overall* ANOVA on the RTs revealed the significant main effects of Sequence, $F(1, 48) = 6.77$, $p = .012$, $\eta_p^2 = .124$, Triplet, $F(1, 48) = 45.90$, $p < .001$, $\eta_p^2 = .489$, and Epoch, $F(5, 240) = 35.91$, $\varepsilon = .612$, $p < .001$, $\eta_p^2 = .428$. The Sequence * Triplet, $F(1, 48) = 22.92$, $p < .001$, $\eta_p^2 = .323$, and Triplet * Epoch, $F(5, 240) = 2.42$, $p = .036$, $\eta_p^2 = .048$, interactions were also significant. In addition, significant Sequence * Group, $F(1, 48) = 69.87$, $p < .001$, $\eta_p^2 = .593$, and Triplet * Group, $F(1, 48) = 4.92$, $p = .031$, $\eta_p^2 = .093$, interactions were revealed. These effects were qualified by the following three-way interactions: The Sequence * Triplet * Epoch interaction was only a trend level, $F(5, 240) = 2.18$, $\varepsilon = .814$, $p = .072$, $\eta_p^2 = .043$, while the Sequence * Epoch * Group, $F(5, 240) = 33.59$, $\varepsilon = .706$, $p < .001$, $\eta_p^2 = .412$, the Triplet * Epoch * Group, $F(5, 240) = 2.31$, $p = .045$, $\eta_p^2 = .046$, and the Sequence * Triplet * Group, $F(1, 48) = 7.96$, $p = .007$, $\eta_p^2 = .142$, interactions were significant. These higher-order effects are detailed below.

The Sequence * Triplet * Epoch interaction tended to indicate that the change of triplet learning over the course of the task differed between the epoch/sequence types. The Sequence * Epoch * Group interaction showed that between-groups differences emerged as a function of first experiencing the structured or the unstructured half (i.e., six epochs) of the task. Particularly, while the Structured-first group became increasingly faster over the structured epochs due to practice and showed similar RTs over the unstructured epochs, this was reversed in the Unstructured-first group, where increasingly faster RTs were observed over the unstructured epochs and similar RTs over the structured epochs (see Fig. 2). This effect suggests that general skill improvements were found in the first half of the task, irrespective of the epoch/sequence type and the presence/absence of transitional probabilities. Relatedly, if

RTs of both groups were considered together, practice effects would not systematically vary as a function of epoch/sequence type because of their reversed order of presentation (non-significant Sequence * Epoch interaction, $F(5, 240) = 1.20$, $\varepsilon = .706$, $p = .313$, $\eta_p^2 = .024$). The overall RTs in the two groups were similar (non-significant main effect of Group, $F(1, 48) = 1.96$, $p = .168$, $\eta_p^2 = .039$) as well as the general skill improvements over the task, irrespective of other experimental effects (non-significant Epoch * Group interaction, $F(5, 240) = 0.47$, $\varepsilon = .612$, $p = .706$, $\eta_p^2 = .010$).

Importantly, the Sequence * Triplet * Group interaction indicated that the difference in triplet learning between the structured and unstructured epochs varied across the groups, which is regarded as the transfer effect (see Fig. 2). This interaction did not reliably vary as a function of practice with the task as shown by the non-significant four-way interaction of Sequence * Triplet * Epoch * Group, $F(5, 240) = 0.61$, $\varepsilon = .814$, $p = .661$, $\eta_p^2 = .012$. According to pair-wise comparisons (see Fig. 3), in the Structured-first group, the extent of triplet learning was similar between the structured and unstructured epochs (6.3 ms vs. 4.6 ms, $p = .171$), while in the Unstructured-first group, it was significantly larger over the structured epochs than over the unstructured epochs (5.9 ms vs. -0.4 ms, $p < .001$). Similarly, the extent of triplet learning did not differ between the groups over the structured epochs (6.3 ms vs. 5.9 ms, $p = .840$) but it was significantly larger in the Structured-first group than in the Unstructured-first group over the unstructured epochs (4.6 ms vs. -0.4 ms, $p < .001$). In the Structured-first group, mean RTs on the high- and low-probability triplets decreased from the structured to the unstructured epochs to a similar extent (high-probability triplets: 392 ms to 383 ms [Diff = 9 ms], $p = .003$; low-probability triplets: 399 ms to 388 ms [Diff = 11 ms], $p < .001$), indicating only general skill improvements across the epoch/sequence types. In the Unstructured-first group, larger RT decrease was found on the high-probability triplets (388 ms to 366 ms [Diff = 22 ms], $p < .001$) than on the low-probability ones (388 ms to 372 ms

[Diff = 16 ms], $p < .001$) from the unstructured to the structured epochs, indicating triplet learning over the structured epochs. The modulating effect of epoch/sequence type on triplet learning was also supported by the Triplet * Epoch * Group interaction, showing that if both task halves with the structured and unstructured epochs were collapsed, the trajectory of triplet learning would differ across the groups. The Sequence * Triplet * Group interaction suggesting the transfer effect is further followed-up below by more detailed analyses.

<PLEASE INSERT FIGURE 2 HERE>

<PLEASE INSERT FIGURE 3 HERE>

Separate analysis of the structured and unstructured epochs

The Triplet by Epoch by Group ANOVA on the RTs related to the *structured* epochs revealed the significant main effects of Triplet, $F(1, 48) = 56.00$, $p < .001$, $\eta_p^2 = .538$, Epoch, $F(5, 240) = 27.54$, $\varepsilon = .700$, $p < .001$, $\eta_p^2 = .365$, and Group, $F(1, 48) = 9.25$, $p = .004$, $\eta_p^2 = .162$. These effects were qualified by the significant Triplet * Epoch, $F(5, 240) = 4.55$, $p = .001$, $\eta_p^2 = .087$, and Epoch * Group, $F(5, 240) = 15.77$, $\varepsilon = .700$, $p < .001$, $\eta_p^2 = .247$, interactions, indicating that triplet learning increased with practice and general skill improvements differed between the groups (see Fig. 2A, D). In regard to the latter effect, while the Structured-first group experienced the structured epochs in the first half of the task ensuring steeper general skill improvements, the Unstructured-first group experienced these epochs after already practicing the stimulus-response mappings for six unstructured epochs, because of which further improvements might have not occurred. Triplet learning and its change over the task did not differ between the groups over the structured epochs as shown by the non-significant Triplet * Group, $F(1, 48) = 0.04$, $p = .840$, $\eta_p^2 = .001$, and Triplet * Epoch * Group interactions, $F(5, 240) = 1.54$, $p = .177$, $\eta_p^2 = .031$. The latter results also indicate that the

order of epoch/sequence types did not induce effects interfering with triplet learning, since the Unstructured-first group, even after a long practice with unstructured sequences, showed similar triplet learning as the Structured-first group.

The same Triplet by Epoch by Group ANOVA on the RTs related to the *unstructured* epochs revealed the significant main effects of Triplet, $F(1, 48) = 10.61$, $p = .002$, $\eta_p^2 = .181$, and Epoch, $F(5, 240) = 13.45$, $\varepsilon = .605$, $p < .001$, $\eta_p^2 = .219$. Importantly, the significant Triplet * Group interaction, $F(1, 48) = 15.23$, $p < .001$, $\eta_p^2 = .241$, showed that triplet learning was larger in the Structured-first group than in the Unstructured-first group (4.6 ms vs. -0.4 ms) over the unstructured epochs. In addition, the trajectory of general skill improvements differed between the groups because of experiencing the unstructured epochs either in the first or the second half of the task (significant Epoch * Group interaction, $F(5, 240) = 16.12$, $\varepsilon = .605$, $p < .001$, $\eta_p^2 = .251$, see Fig. 2B, C). The groups did not differ in overall RT and in the change of triplet learning over the unstructured epochs as shown by the non-significant Group main effect, $F(1, 48) = 0.09$, $p = .764$, $\eta_p^2 = .002$, and Triplet * Epoch * Group interaction, $F(5, 240) = 1.33$, $p = .251$, $\eta_p^2 = .027$ (the Triplet * Epoch interaction was non-significant, $F(5, 240) = 0.24$, $p = .945$, $\eta_p^2 = .005$, either).

Performance in the Structured-first group

Considering only the *Structured-first group*, the Sequence by Triplet by Epoch ANOVA revealed the main effects of Sequence, $F(1, 24) = 14.31$, $p = .001$, $\eta_p^2 = .374$, Triplet, $F(1, 24) = 32.62$, $p < .001$, $\eta_p^2 = .576$, and Epoch, $F(5, 120) = 17.59$, $\varepsilon = .665$, $p < .001$, $\eta_p^2 = .423$. Accordingly, this group showed triplet knowledge throughout the task, which, importantly, did not change in time (non-significant Triplet * Epoch interaction, $F(5, 120) = 0.87$, $p = .502$, $\eta_p^2 = .035$) and did not differ between the structured and unstructured epochs (non-significant Sequence * Triplet, $F(1, 24) = 1.54$, $p = .226$, $\eta_p^2 = .060$, and Sequence * Triplet * Epoch interactions, $F(5, 120) = 1.52$, $\varepsilon = .669$, $p = .211$, $\eta_p^2 = .060$). As participants in this group

started the task with its structured half, general skill improvements over the structured epochs and a consistent speed of responding over the unstructured epochs were also found (significant Sequence * Epoch interaction, $F(5, 120) = 25.28$, $\varepsilon = .693$, $p < .001$, $\eta_p^2 = .513$, see Fig. 2A, B).

The Triplet by Epoch ANOVA conducted on the RTs measured separately in the *structured epochs* in the *Structured-first group* revealed significant triplet learning and general skill improvements (significant main effects of Triplet, $F(1, 24) = 19.58$, $p < .001$, $\eta_p^2 = .449$, and Epoch, $F(5, 120) = 32.85$, $\varepsilon = .611$, $p < .001$, $\eta_p^2 = .578$; non-significant Triplet * Epoch interaction, $F(5, 120) = 1.59$, $p = .168$, $\eta_p^2 = .062$, see Fig. 2A). In the *unstructured epochs*, triplet knowledge was also observed (significant main effect of Triplet, $F(1, 24) = 31.16$, $p < .001$, $\eta_p^2 = .565$; the main effect of Epoch, $F(5, 120) = 1.96$, $\varepsilon = .685$, $p = .118$, $\eta_p^2 = .076$, and the Triplet * Epoch interaction, $F(5, 120) = 0.65$, $p = .664$, $\eta_p^2 = .026$, were not significant, see Fig. 2B), corroborating the above-reported significant Triplet main effect from the Sequence by Triplet by Epoch ANOVA performed in this group.

Performance in the Unstructured-first group

Considering only the *Unstructured-first group*, the Sequence by Triplet by Epoch ANOVA revealed the main effects of Sequence, $F(1, 24) = 71.32$, $p < .001$, $\eta_p^2 = .748$, Triplet, $F(1, 24) = 13.65$, $p = .001$, $\eta_p^2 = .363$, and Epoch, $F(5, 120) = 18.70$, $\varepsilon = .544$, $p < .001$, $\eta_p^2 = .438$. As participants in this group started the task with its unstructured half, general skill improvements due to practice over the unstructured epochs and a consistent speed of responding over the structured epochs were found (significant Sequence * Epoch interaction, $F(5, 120) = 11.50$, $\varepsilon = .689$, $p < .001$, $\eta_p^2 = .324$). Triplet learning increased as the task progressed (significant Triplet * Epoch interaction, $F(5, 120) = 4.55$, $\varepsilon = .768$, $p = .002$, $\eta_p^2 = .159$). However, this effect should be interpreted with the consideration of the significant Sequence * Triplet interaction, $F(1, 24) = 38.79$, $p < .001$, $\eta_p^2 = .618$, indicating that triplet

learning was virtually zero on the unstructured epochs (-0.4 ms) but it gradually emerged across the structured epochs (5.9 ms). Nevertheless, the three-way Sequence * Triplet * Epoch interaction was not significant, $F(5, 120) = 1.22$, $p = .302$, $\eta_p^2 = .048$ (see Fig. 2C, D).

The Triplet by Epoch ANOVA conducted on the RTs measured separately in the *unstructured epochs* in the *Unstructured-first group* revealed significant general skill improvements (significant main effect of Epoch, $F(5, 120) = 20.41$, $\varepsilon = .518$, $p < .001$, $\eta_p^2 = .460$), but triplet learning could not be reliably detected (non-significant main effect of Triplet, $F(1, 24) = 0.18$, $p = .678$, $\eta_p^2 = .007$, and Triplet * Epoch interaction, $F(5, 120) = 0.90$, $p = .481$, $\eta_p^2 = .036$, see Fig. 2C). Over the *structured epochs*, triplet learning increased (significant main effect of Triplet, $F(1, 24) = 53.92$, $p < .001$, $\eta_p^2 = .692$, and Triplet * Epoch interaction $F(5, 120) = 6.48$, $p < .001$, $\eta_p^2 = .213$) from 1 ms in $epoch_1$ to 12 ms in $epoch_6$ (the difference between high- and low-probability triplets remained significant from $epoch_3$ to $epoch_6$, all *p*s $\leq .015$, see Fig. 2D). The observed effects on the separate halves of the task corroborate the above-reported Triplet * Epoch and Sequence * Triplet interactions from the Sequence by Triplet by Epoch ANOVA performed in this group.

Discussion

This study investigated whether the implicitly acquired knowledge on a 2nd order transitional probability structure influences the processing of unpredictable transitional probabilities throughout a learning situation. To this end, the changes of RTs to more probable and less probable short-range transitional probabilities (triplets) embedded in a stimulus sequence were tracked. The stimulus sequence changed over the experimental task since an alternating regularity supporting the extraction of triplets was present in one half of the task blocks and absent in the other half, without explicitly denoting this change at the surface level. In line with our assumptions, while the participant group completing first the structured half of the

task showed triplet learning across both the structured and unstructured blocks, the participant group completing first the unstructured half showed triplet learning only over the structured blocks and not over the unstructured blocks. Based on this between-groups difference, we suggest that the acquired implicit knowledge on the short-range transitional probabilities has been transferred or generalized across the learning situations. Importantly, this prior knowledge has been transferred to a stimulus stream with an equal transitional probability structure, consisting only of uniformly distributed stimulus chunks that otherwise could not be extracted.

Previously, in the study of Gaissmaier and Schooler (2008), participants completed a binary choice task where the transition between the no pattern (no serial dependence in the sequence) and pattern half (repeating deterministic sequence) was clearly denoted. The study used a within-subjects design and the order of task halves was either counterbalanced or fixed across participants. Results indicated that participants who were more prone to search for patterns in the no pattern half of the task showed higher accuracy in the pattern half as compared with those participants who were less prone to follow any complex search strategy. This is partially in line with the present findings since participants completing first the structured blocks similarly *perceived* the relations of stimuli in the unstructured blocks as in the structured blocks. An important difference between the studies is, however, that the support of the alternating regularity that could have triggered the later applied implicit search strategies was present during the first half of the task only in one of the groups. In addition, participants did not know that they were in a learning situation and they did not have information on whether the sequence of stimuli was random or followed any underlying pattern. These between-study differences nevertheless suggest that the perception of participants could have been influenced by the *primarily experienced* transitional probability structure. In other words, because of the task environment, participants first training on the

structured sequence might have worked up a tendency towards pattern detection, which could have resulted in forming implicit expectations about the upcoming stimuli. Meanwhile, some of the participants (in both groups) might have inherently possessed a baseline tendency towards pattern detection but this was not investigated here.

The constraining effect of the primarily acquired statistical structure was also shown by the study of Gebhart et al. (2009), where, like in the present experimental design, no explicit information was provided about the change in structure during the task or the second structure was not presented for a longer duration. Regarding the ASRT task, it has already been demonstrated that the acquired implicit knowledge on the transitional probability structure remains *stable* over longer time periods such as one week (Nemeth & Janacsek, 2011) or even one year (Romano, Howard, & Howard, 2010). Moreover, this knowledge is resistant to short periods of *interfering* sequences inserted into the course of learning not only after 24 hours but also after one year has elapsed (Kóbor et al., 2017).

In the Kóbor et al. (2017) study, the alternating regularity guiding the interference sequence partially overlapped with that of the primarily practiced, original sequence. This changed the frequency of the originally high- and low-probability triplets (for instance, 75% of the high-probability triplets became low-probability ones), which resulted only in a modest degree of triplet learning on the interference sequence. Meanwhile, in the present study, the extent of triplet learning was comparable across the structured and unstructured blocks in participants first completing the structured sequence (see Fig. 2A–B and Fig. 3). This might originate from the fact that the unique triplets occurred with equal probability (25%) in the unstructured sequence, yielding a simpler probability distribution than that of the statistical structure used in the study of Kóbor et al. (2017). In addition, the absence of the alternating regularity might have reduced or even eliminated any interference by not supporting the extraction of the changed transitional probabilities (as observed in the Kóbor et al. (2017)

study). Instead, after short experience with the unstructured sequence, participants first completing the structured sequence might have implicitly identified the features common (i.e., triplets) across the two task halves, and this might have supported the generalization of the acquired transitional probability structure (cf., Robertson, 2018). This way, the transfer effect could also be interpreted as the recall and implementation of the probabilistic structure. In relation to the ASRT task, the former and the present results altogether suggest that knowledge at the level of short-range transitional probabilities are acquired and the formed probabilistic representations are *robust*.

Establishing robust representations on the experienced statistical structure was found to be advantageous in relation to various perceptual and cognitive processes (cf. Gebhart et al., 2009; Todd et al., 2011). In the ASRT task, not only processes underlying implicit statistical learning are involved but also *serial decision making* takes place during task solving (Török, Janacsek, Nagy, Orbán, & Nemeth, 2017). Considering the latter process, it was previously shown that information integrated over the *recent history of stimuli* influenced elemental visual perception (Fischer & Whitney, 2014) and post-perceptual decisions (Fritsche, Mostert, & de Lange, 2017), which, by promoting the stability of perceptual experience, possibly supported adaptation to the ever-changing visual input. The short-range serial dependence already observed in perceptual decisions inherently characterizes decisions in the ASRT task due to the acquisition of the transitional probability structure. Beyond short-range serial dependence, long-range ordered dependence biasing upcoming decisions has been demonstrated, as well (J. H. Howard, Jr. & Howard, 1997; Kóbor et al., 2018; Nemeth, Janacsek, & Fiser, 2013; Simor et al., 2019). Furthermore, by showing triplet learning across the task halves, the present findings indicate even the *transfer of serial dependence*. Overall, the obtained transfer effect bolsters the idea that the robustness of the acquired representations might provide fast and automatic adaptation to the stimulus stream.

Contrary to robustness, *instability* of memory representations was found to increase their susceptibility to interference and thereby promoted learning transfer, at least between different memory tasks (Mosha & Robertson, 2016). Although the notion of memory instability might not easily account for the observed findings, it has been suggested that, via generalization, successful adaptation to a range of tasks requires the *forgetting* or weakening of some specific features of the acquired representations (Robertson, 2018). In the present case, this would have been the forgetting of the initial triplet frequency information and/or the structure of the alternating regularity when starting the other task half. Particularly, it is probable that participants first completing the structured blocks became gradually sensitive to the altered frequency of unique triplets. In support of this idea, in Figure 2B, it is observable that RTs to high- and low-probability triplets are approaching one another in the second half of the unstructured sequence ($epochs_{4,5,6}$), decreasing the extent of triplet knowledge. In particular, the extent of triplet knowledge averaged over the last two structured epochs is significantly larger than over the last two unstructured epochs (7.7 ms vs. 2.8 ms, $t(24) = 2.46$, $p = .022$). It is plausible to assume that after having had performed even more unstructured blocks, these participants would have learned that the originally high-probability triplets no longer occurred with higher frequency than the originally low-probability ones, and, therefore, the initial triplet frequency information would have been forgotten or “unlearned”. This could be in accordance with the findings of Szegedi-Hallgató et al. (2017), indicating the coexistence of the previously and the recently acquired implicit knowledge on the changed statistical structure in the ASRT task. Considering the processes underlying these RT changes, instead of “unlearning”, the forming of new memories or the reconsolidation of the previously created ones would also be possible (Chandler & Gass, 2013). Overall, either concept supports the interpretation that in the present experiment, stimulus processing was determined by a prior internal model on the transitional probability structure that changed slowly with

accumulating experiences about the ongoing stimulus context (cf. Daw et al., 2011; Griffiths et al., 2008; Shohamy & Daw, 2015). Future studies might investigate the effect of long-term practice with different structured sequences and whether a flexible set of representations applied in and transferred across conditions can be established.

Beyond triplet frequency information, the importance of the *alternating regularity* is highlighted by the performance of the group first completing the unstructured sequence. Although the short-range transitional probabilities over the unstructured blocks were presented to this group in the same manner as to the other one, the order of task halves did matter (cf. Bischoff-Grethe, Martin, Mao, & Berns, 2001). The use of the between-groups design assured the appropriate testing of the learning transfer. In addition, it implied that if no rule determined the stimulus presentation when primary experiences were gathered, from a uniform distribution of stimuli (i.e., triplets appearing with equal probability), transitional probabilities could not be extracted. Nevertheless, we cannot exclude the possibility that even without the alternating regularity, triplet learning would occur if the probabilities of the high- and low-probability triplets were changed otherwise (e.g., keeping the original probability of occurrence of high- [62.5%] and low-probability [37.5%] triplets). To clarify the role of the alternating regularity in task performance, future studies should more directly investigate how and when the coding of stimuli according to their position in the probabilistic sequence takes place.

In terms of methodological issues, as the study of Reed and Johnson (1994) emphasizes, to properly test the acquisition of complex statistical structures, one should use training and transfer sequences differing only in 2nd order transitional probabilities instead of using a random transfer sequence. Therefore, we deliberately avoided the use of *fully random* sequences; and, instead, we applied “equal probability” unstructured sequences, which were more controlled than the former ones. Since picking up transitional and distributional statistics

from the ongoing sensory environment is considered as an automatic and effortless human capacity (Fiser & Aslin, 2002; Orbán et al., 2008; Siegelman et al., 2017), individuals should be tuned to implicitly perceive various temporal contingencies in the sensory input instead of regarding it as a stream of random stimuli (cf. Warren et al., 2018). This would have been undesirable in the present experiment. In addition, previous research also demonstrated that participants would less likely search for underlying patterns if a sequence, compared with another, was subjectively perceived as more random, but, according to objective measures, was more structured (Wolford, Newman, Miller, & Wig, 2004). Since triplet learning did not occur over the unstructured blocks in participants first completing these blocks, it is conceivable that the unstructured sequence, at least for the first experience, was perceived as truly random, however, it was, indeed, characterized by an “equal probability” structure.

In sum, the present experiment provided evidence for the implicit memory transfer of a complex transitional probability structure. From certain aspects, the obtained findings could contribute to the understanding of how implicit prior knowledge influences further perceptual and cognitive processing. Particularly, our results indicate that, due to the prior settings of perception and thereby the robustness of the acquired internal models, a stimulus environment lacking a predictable transitional probability structure is automatically processed according to the previously observed statistical structure. This means that the ongoing stimulus context is perceived in a model-based manner. Moreover, the extraction and transfer of short-range serial dependence seems to be present not only at the level of visual perception but also at the level of serial decision making. Finally, this study also highlights the importance of carefully constructing the underlying structure of training and transfer sequences in the investigation of statistical-sequence learning.

Acknowledgement

This research was supported by the National Brain Research Program (project 2017-1.2.1-NKP-2017-00002, PI: D. N.); the Hungarian Scientific Research Fund (OTKA FK 124412, PI: A. K., OTKA PD 124148, PI: K. J., OTKA K 128016, PI: D. N.); the IDEXLYON Fellowship (to D. N.); and the János Bolyai Research Scholarship of the Hungarian Academy of Sciences (to A. K. and K. J.). The authors thank Balázs Török for providing custom-written scripts that generated the unstructured sequences and Borbála German for helping in data acquisition.

Declarations of interest

None.

References

Armstrong, B. C., Frost, R., & Christiansen, M. H. (2017). The long road of statistical learning research: Past, present and future. *Philosophical Transactions of the Royal Society B: Biological Sciences, 372*(1711). doi:10.1098/rstb.2016.0047

Aslin, R. N. (2017). Statistical learning: A powerful mechanism that operates by mere exposure. *Wiley Interdisciplinary Reviews: Cognitive Science, 8*(1-2), e1373. doi:10.1002/wcs.1373

Bischoff-Grethe, A., Martin, M., Mao, H., & Berns, G. S. (2001). The context of uncertainty modulates the subcortical response to predictability. *Journal of Cognitive Neuroscience, 13*(7), 986-993. doi:10.1162/089892901753165881

Brainard, D., H. (1997). The psychophysics toolbox. *Spatial Vision, 10*(4), 433-436. doi:10.1163/156856897X00357

Chandler, L., & Gass, J. (2013). The plasticity of extinction: Contribution of the prefrontal cortex in treating addiction through inhibitory learning. *Frontiers in Psychiatry, 4*(46). doi:10.3389/fpsyt.2013.00046

Christiansen, M. H. (2018). Implicit statistical learning: A tale of two literatures. *Topics in Cognitive Science*. doi:10.1111/tops.12332

Conway, C. M., & Christiansen, M. H. (2001). Sequential learning in non-human primates. *Trends in Cognitive Sciences, 5*(12), 539-546. doi:10.1016/S1364-6613(00)01800-3

Daltrozzo, J., & Conway, C. M. (2014). Neurocognitive mechanisms of statistical-sequential learning: What do event-related potentials tell us? *Frontiers in Human Neuroscience, 8*(437). doi:10.3389/fnhum.2014.00437

Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., & Dolan, R. J. (2011). Model-based influences on humans' choices and striatal prediction errors. *Neuron, 69*(6), 1204-1215. doi:10.1016/j.neuron.2011.02.027

Dragovic, M. (2004a). Categorization and validation of handedness using latent class analysis. *Acta Neuropsychiatrica, 16*(4), 212-218. doi:10.1111/j.0924-2708.2004.00087.x

Dragovic, M. (2004b). Towards an improved measure of the Edinburgh Handedness Inventory: A one-factor congeneric measurement model using confirmatory factor analysis. *Laterality: Asymmetries of Body, Brain and Cognition, 9*(4), 411-419. doi:10.1080/13576500342000248

Feher da Silva, C., & Baldo, M. V. C. (2012). A simple artificial life model explains irrational behavior in human decision-making. *PLoS ONE, 7*(5), e34371. doi:10.1371/journal.pone.0034371

Fischer, J., & Whitney, D. (2014). Serial dependence in visual perception. *Nature Neuroscience, 17*, 738. doi:10.1038/nn.3689

Fiser, J., & Aslin, R. N. (2002). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences, 99*(24), 15822-15826. doi:10.1073/pnas.232472899

Friston, K. (2005). A theory of cortical responses. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 360*(1456), 815-836. doi:10.1098/rstb.2005.1622

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience, 11*(2), 127-138. doi:10.1038/nrn2787

Friston, K. J., Stephan, K. E., Montague, R., & Dolan, R. J. (2014). Computational psychiatry: The brain as a phantastic organ. *The Lancet Psychiatry, 1*(2), 148-158. doi:10.1016/S2215-0366(14)70275-5

Fritsche, M., Mostert, P., & de Lange, F. P. (2017). Opposite effects of recent history on perception and decision. *Current Biology, 27*(4), 590-595. doi:10.1016/j.cub.2017.01.006

Frost, R. L. A., & Monaghan, P. (2016). Simultaneous segmentation and generalisation of non-adjacent dependencies from continuous speech. *Cognition, 147*, 70-74. doi:10.1016/j.cognition.2015.11.010

Gaissmaier, W., & Schooler, L. J. (2008). The smart potential behind probability matching. *Cognition, 109*(3), 416-422. doi:10.1016/j.cognition.2008.09.007

Gebhart, A. L., Aslin, R. N., & Newport, E. L. (2009). Changing structures in midstream: Learning along the statistical garden path. *Cognitive Science, 33*(6), 1087-1116. doi:10.1111/j.1551-6709.2009.01041.x

Greenhouse, S., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika, 24*(2), 95-112. doi:10.1007/bf02289823

Griffiths, T. L., Kemp, C., & Tenenbaum, J. B. (2008). Bayesian models of cognition. In R. Sun (Ed.), *Cambridge handbook of computational cognitive modeling*: Cambridge University Press.

Hahn, U., & Warren, P. A. (2009). Perceptions of randomness: Why three heads are better than four. *Psychological Review, 116*(2), 454-461. doi:10.1037/a0015241

Hallgató, E., Győri-Dani, D., Pekár, J., Janacsek, K., & Nemeth, D. (2013). The differential consolidation of perceptual and motor learning in skill acquisition. *Cortex, 49*(4), 1073-1081. doi:10.1016/j.cortex.2012.01.002

Honbolygó, F., & Csépe, V. (2013). Saliency or template? ERP evidence for long-term representation of word stress. *International Journal of Psychophysiology, 87*(2), 165-172. doi:10.1016/j.ijpsycho.2012.12.005

Howard, D. V., Howard, J. H., Jr., 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. doi:10.1037/0882-7974.19.1.79

Howard, J. H., Jr., & Howard, D. V. (1997). Age differences in implicit learning of higher order dependencies in serial patterns. *Psychology and Aging, 12*(4), 634-656. doi:10.1037/0882-7974.12.4.634

James, G., & Koehler, D. J. (2011). Banking on a bad bet. Probability matching in risky choice is linked to expectation generation. *Psychological Science, 22*(6), 707-711. doi:10.1177/0956797611407933

Janacsek, K., Ambrus, G. G., Paulus, W., Antal, A., & Nemeth, D. (2015). Right hemisphere advantage in statistical learning: Evidence from a probabilistic sequence learning task. *Brain Stimulation, 8*(2), 277-282. doi:dx.doi.org/10.1016/j.brs.2014.11.008

Kaufman, S. B., Deyoung, C. G., Gray, J. R., Jimenez, L., Brown, J., & Mackintosh, N. (2010). Implicit learning as an ability. *Cognition, 116*(3), 321-340. doi:10.1016/j.cognition.2010.05.011

Kóbor, A., Janacsek, K., Takács, Á., & Nemeth, D. (2017). Statistical learning leads to persistent memory: Evidence for one-year consolidation. *Scientific Reports, 7*(1), 760. doi:10.1038/s41598-017-00807-3

Kóbor, A., Takács, Á., Kardos, Z., Janacsek, K., Horváth, K., Csépe, V., & Nemeth, D. (2018). ERPs differentiate the sensitivity to statistical probabilities and the learning of sequential structures during procedural learning. *Biological Psychology, 135*, 180-193. doi:10.1016/j.biopsycho.2018.04.001

Maheu, M., Dehaene, S., & Meyniel, F. (2019). Brain signatures of a multiscale process of sequence learning in humans. *eLife, 8*, e41541. doi:10.7554/eLife.41541

Malassis, R., Rey, A., & Fagot, J. (2018). Non□adjacent dependencies processing in human and non□human primates. *Cognitive Science*. doi:10.1111/cogs.12617

Meyniel, F., Maheu, M., & Dehaene, S. (2016). Human inferences about sequences: A minimal transition probability model. *PLoS Computational Biology, 12*(12), e1005260-e1005260. doi:10.1371/journal.pcbi.1005260

Mosha, N., & Robertson, E. M. (2016). Unstable memories create a high-level representation that enables learning transfer. *Current Biology, 26*(1), 100-105. doi:10.1016/j.cub.2015.11.035

Mueller, J. L., Milne, A., & Männel, C. (2018). Non-adjacent auditory sequence learning across development and primate species. *Current Opinion in Behavioral Sciences, 21*, 112-119. doi:10.1016/j.cobeha.2018.04.002

Mullens, D., Woodley, J., Whitson, L., Provost, A., Heathcote, A., Winkler, I., & Todd, J. (2014). Altering the primacy bias—how does a prior task affect mismatch negativity? *Psychophysiology, 51*(5), 437-445. doi:10.1111/psyp.12190

Nemeth, D., & Janacsek, K. (2011). The dynamics of implicit skill consolidation in young and elderly adults. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences, 66*(1), 15-22. doi:10.1093/geronb/gbq063

Nemeth, D., Janacsek, K., & Fiser, J. (2013). Age-dependent and coordinated shift in performance between implicit and explicit skill learning. *Frontiers in Computational Neuroscience, 7*, 147. doi:10.3389/fncom.2013.00147

Nemeth, D., Janacsek, K., Londe, Z., Ullman, M. T., Howard, D. V., & Howard, J. H., Jr. (2010). Sleep has no critical role in implicit motor sequence learning in young and old adults. *Experimental Brain Research, 201*(2), 351-358. doi:10.1007/s00221-009-2024-x

Nemeth, D., Janacsek, K., Polner, B., & Kovacs, Z. A. (2013). Boosting human learning by hypnosis. *Cerebral Cortex, 23*(4), 801-805. doi:10.1093/cercor/bhs068

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh Inventory. *Neuropsychologia, 9*(1), 97-113. doi:10.1016/0028-3932(71)90067-4

Orbán, G., Fiser, J., Aslin, R. N., & Lengyel, M. (2008). Bayesian learning of visual chunks by human observers. *Proceedings of the National Academy of Sciences, 105*(7), 2745-2750. doi:10.1073/pnas.0708424105

Pelli, D., G. (1997). The videotoolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*(4), 437-442. doi:10.1163/156856897X00366

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), 585-594. doi:10.1037/0278-7393.20.3.585

Rey, A., Minier, L., Malassis, R., Bogaerts, L., & Fagot, J. (2018). Regularity extraction across species: Associative learning mechanisms shared by human and non□human primates. *Topics in Cognitive Science*. doi:10.1111/tops.12343

Robertson, E. M. (2018). Memory instability as a gateway to generalization. *PLoS Biology, 16*(3), e2004633. doi:10.1371/journal.pbio.2004633

Romano, J. C., Howard, J. H., Jr., & Howard, D. V. (2010). One-year retention of general and sequence-specific skills in a probabilistic, serial reaction time task. *Memory, 18*(4), 427-441. doi:10.1080/09658211003742680

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

Shohamy, D., & Daw, N. D. (2015). Integrating memories to guide decisions. *Current Opinion in Behavioral Sciences, 5*, 85-90. doi:10.1016/j.cobeha.2015.08.010

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). doi:10.1098/rstb.2016.0059

Simor, P., Zavecz, Z., Horváth, K., Éltető, N., Török, C., Pesthy, O., . . . Nemeth, D. (2019). Deconstructing procedural memory: Different learning trajectories and consolidation of sequence and statistical learning. *Frontiers in Psychology, 9*(2708). doi:10.3389/fpsyg.2018.02708

Song, S., Howard, J. H., Jr., & Howard, D. V. (2007). Sleep does not benefit probabilistic motor sequence learning. *The Journal of Neuroscience, 27*(46), 12475-12483. doi:10.1523/jneurosci.2062-07.2007

Sun, Y., O'Reilly, R. C., Bhattacharyya, R., Smith, J. W., Liu, X., & Wang, H. (2015). Latent structure in random sequences drives neural learning toward a rational bias.

Proceedings of the National Academy of Sciences of the United States of America, 112(12), 3788-3792. doi:10.1073/pnas.1422036112

Sun, Y., & Wang, H. (2010). Gambler's fallacy, hot hand belief, and the time of patterns. *Judgment and Decision Making, 5*(2), 124-132.

Szegedi-Hallgató, E., Janacsek, K., Vékony, T., Tasi, L. A., Kerepes, L., Hompoth, E. A., . . . Németh, D. (2017). Explicit instructions and consolidation promote rewiring of automatic behaviors in the human mind. *Scientific Reports, 7*(1), 4365. doi:10.1038/s41598-017-04500-3

Takács, Á., Kóbor, A., Chezan, J., Éltető, N., Tárnok, Z., Nemeth, D., . . . Janacsek, K. (2018). Is procedural memory enhanced in Tourette syndrome? Evidence from a sequence learning task. *Cortex, 100*, 84-94. doi:10.1016/j.cortex.2017.08.037

Takács, Á., Shilon, Y., Janacsek, K., Kóbor, A., Tremblay, A., Nemeth, D., & Ullman, M. T. (2017). Procedural learning in Tourette syndrome, ADHD, and comorbid Tourette-ADHD: Evidence from a probabilistic sequence learning task. *Brain and Cognition, 117*, 33-40. doi:10.1016/j.bandc.2017.06.009

Todd, J., Provost, A., & Cooper, G. (2011). Lasting first impressions: A conservative bias in automatic filters of the acoustic environment. *Neuropsychologia, 49*(12), 3399-3405. doi:10.1016/j.neuropsychologia.2011.08.016

Tóth, B., Janacsek, K., Takács, Á., Kóbor, A., Zavecz, Z., & Nemeth, D. (2017). Dynamics of EEG functional connectivity during statistical learning. *Neurobiology of Learning and Memory, 144*, 216-229. doi:10.1016/j.nlm.2017.07.015

Török, B., Janacsek, K., Nagy, D. G., Orbán, G., & Nemeth, D. (2017). Measuring and filtering reactive inhibition is essential for assessing serial decision making and learning. *Journal of Experimental Psychology: General, 146*(4), 529-542. doi:10.1037/xge0000288

Virag, M., Janacsek, K., Horvath, A., Bujdoso, Z., Fabo, D., & Nemeth, D. (2015). Competition between frontal lobe functions and implicit sequence learning: Evidence from the long-term effects of alcohol. *Experimental Brain Research, 233*(7), 2081-2089. doi:10.1007/s00221-015-4279-8

Warren, P. A., Gostoli, U., Farmer, G. D., El-Deredy, W., & Hahn, U. (2018). A re-examination of "bias" in human randomness perception. *Journal of Experimental Psychology. Human Perception and Performance, 44*(5), 663-680. doi:10.1037/xhp0000462

Wolford, G., Newman, S. E., Miller, M. B., & Wig, G. S. (2004). Searching for patterns in random sequences. *Canadian Journal of Experimental Psychology, 58*(4), 221-228.

Tables

Table 1. Descriptive data and performance on neuropsychological tests in the two groups.

	Structured-first group *n* = 25 *M*(*SD*)	Unstructured-first group *n* = 25 *M*(*SD*)	Between-groups difference $t/U/\chi^2$
Gender [Male/Female]	8/17	8/17	1.00
Age [years]	21.5 (2.6)	20.8 (1.5)	274.50[a]
Education [years]	14.6 (2.1)	14.3 (1.6)	0.69
Handedness [LQ]	53.5 (49.3)	65.8 (28.7)	330.50[a]
Wisconsin Card Sorting Task [perseverative error percentage]	11.67 (3.81)	10.41 (2.02)	1.46[b]
Corsi blocks task [visuospatial short-term memory span; range: 3–9]	5.13 (0.54)	5.04 (0.68)	0.49
Counting span task [working memory span; range: 2–6]	3.80 (0.86)	4.08 (0.85)	-1.16
Go/No-Go task [discriminability index: hit rate minus false alarm rate]	.72 (.14)	.71 (.15)	0.40

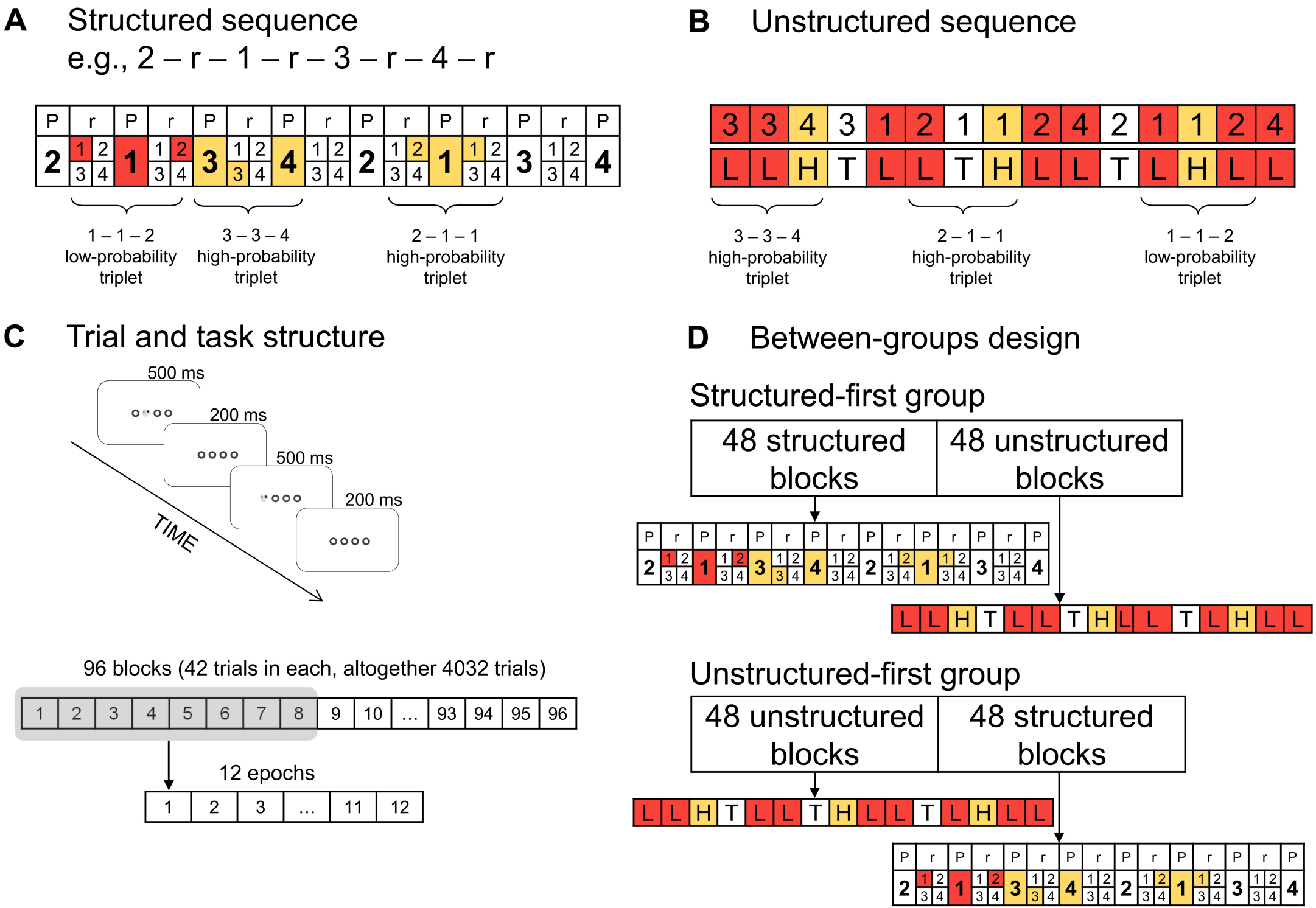
Note. The two groups *did not differ* in any of the dependent variables and all participants performed in the normal range on the neuropsychological tests. [a]In the case of violating the assumption of normality, the Mann-Whitney U test was performed, and the *U* statistic is provided. [b]In case of violating the assumption of homogeneity of variances, the robust Welch test of equality of means was performed, and the *t* statistic is provided. Handedness was assessed with the Edinburgh Handedness Inventory revised version (Dragovic, 2004a, 2004b; Oldfield, 1971); LQ = Laterality Quotient, −100 means complete left-handedness, 100 means complete right-handedness.

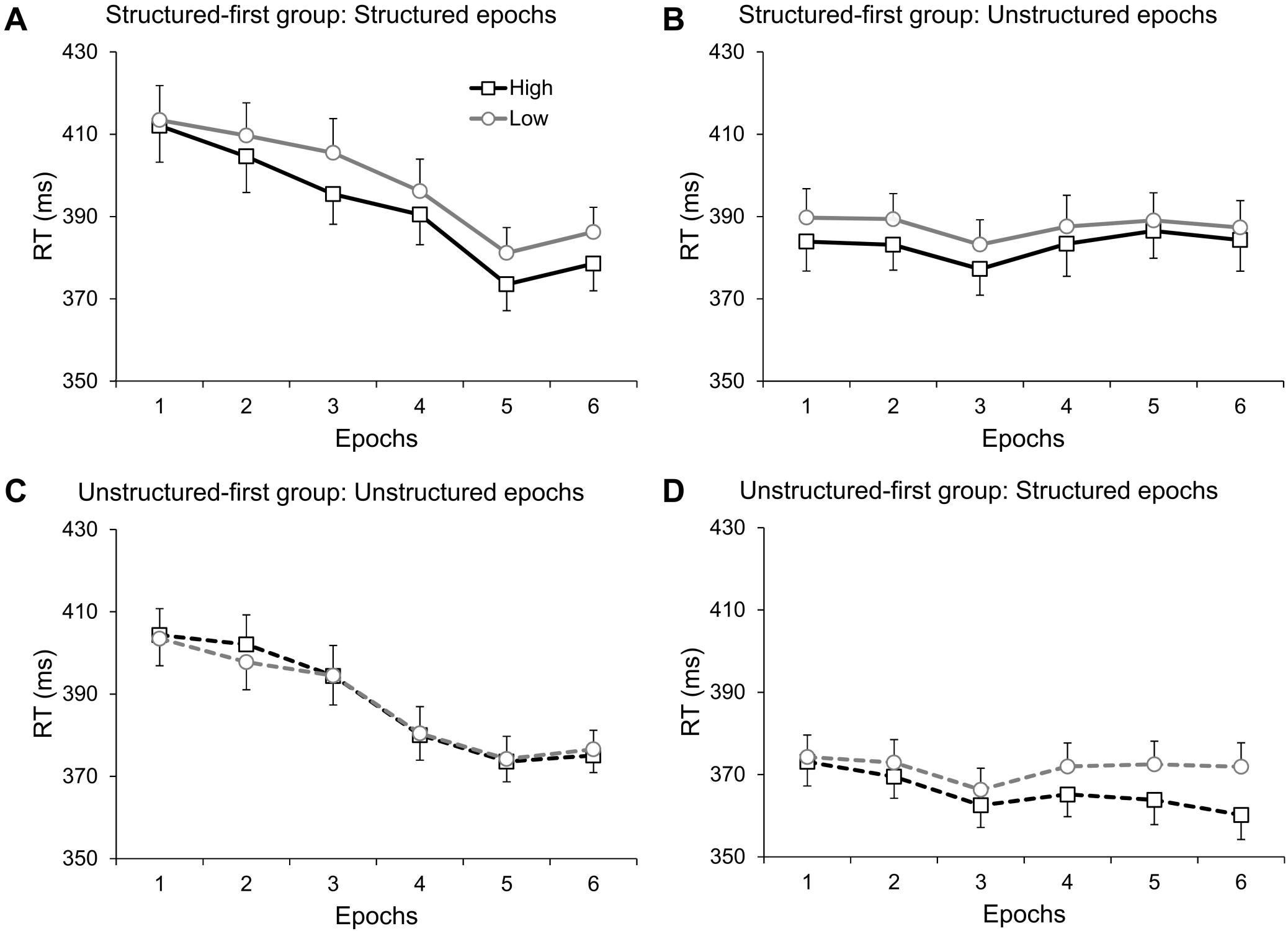
Figure captions

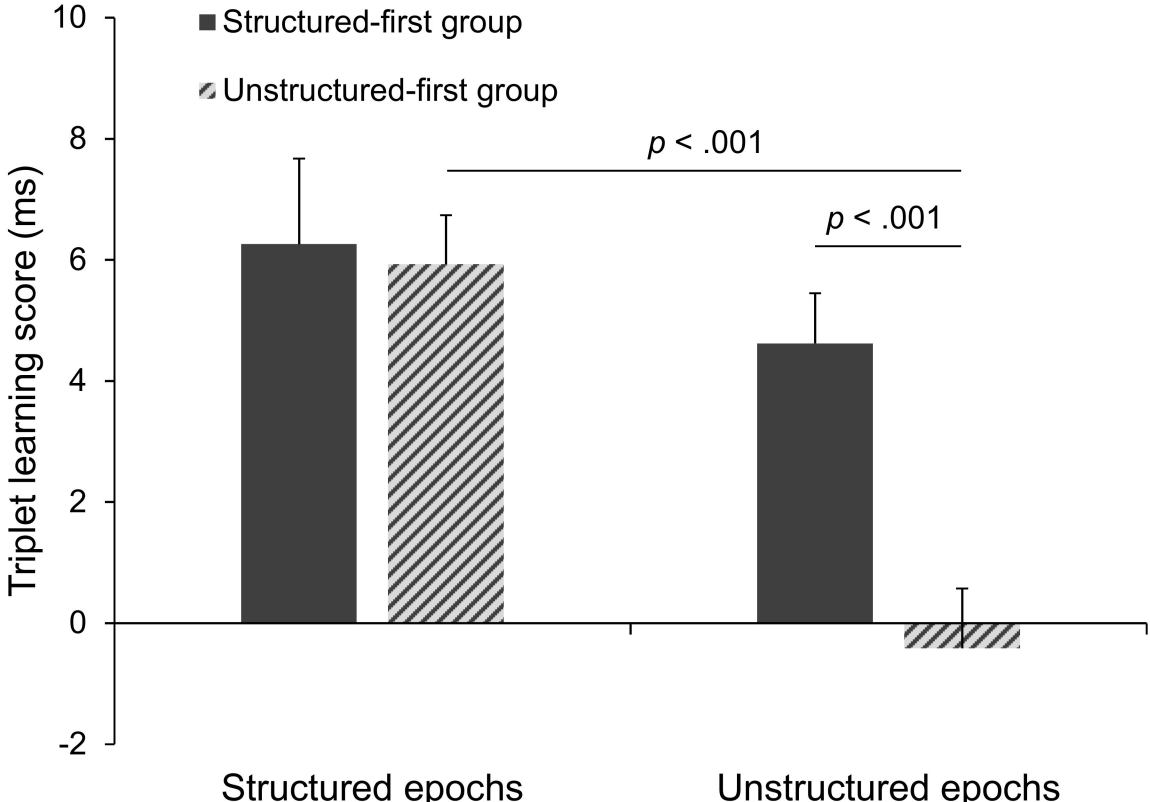
Figure 1. Design of the experiment. (A) The presentation of stimuli in the structured ASRT sequence followed an eight-element regularity, within which pattern (P) and random (r) elements alternated with one another. Numbers denote the four different stimulus positions on the screen. The alternating regularity made some runs of three consecutive trials (triplets) more probable than others. High-probability triplets are denoted with gold shading and low-probability triplets are denoted with coral shading. (B) From the unstructured sequence, the alternating regularity was omitted, but the same unique triplets as in the structured sequence appeared with equal probability. Gold shading (upper row) and capital letter "H" (lower row) denote the third element of high-probability triplets, coral shading and "L" denote the third element of low-probability triplets, while white shading and "T" as "trill" denote the third element of some of those low-probability triplets that were eliminated from the analyses (see section Statistical analysis). Numbers denote the four different stimulus positions on the screen. Note that each stimulus (trial) is categorized as either the third element of a high- or a low-probability triplet in both sequences. For a given participant, at the level of unique triplets, the high- and low-probability triplets are the same in the structured and unstructured sequences. (C) In this version of the task, a stimulus appeared in one of four horizontally arranged empty circles on the screen in every 700 ms. Participants had to respond with one of the four response keys that corresponded to the position of the stimulus. They completed altogether 96 blocks, and eight-block-long units of the task were collapsed into larger time bins labeled as epochs. (D) While the Structured-first group (n = 25) completed 48 structured blocks followed by 48 unstructured blocks, the Unstructured-first group (n = 25) completed 48 unstructured blocks followed by 48 structured blocks.

Figure 2. Temporal dynamics of triplet learning across groups and epoch/sequence types. Group-average RTs (A–B: Structured-first group; C–D: Unstructured-first group) for correct responses as a function of time bin (epochs 1–6) and triplet type (high- vs. low-probability triplets) are presented in the structured (A, D) and unstructured (B, C) epochs. Error bars denote standard error of mean.

Figure 3. Transfer of the acquired implicit knowledge. Group-average overall (mean) triplet learning scores (RTs to low- minus RTs to high-probability triplets) are presented in the Structured-first and Unstructured-first groups for the structured and unstructured epochs, respectively. Error bars denote standard error of mean.


A
Structured sequence
e.g., 2 – r – 1 – r – 3 – r – 4 – r
1 – 1 – 2
low-probability
triplet
3 – 3 – 4
high-probability
triplet
2 – 1 – 1
high-probability
triplet
B
Unstructured sequence
3 – 3 – 4
high-probability
triplet
2 – 1 – 1
high-probability
triplet
1 – 1 – 2
low-probability
triplet
C
Trial and task structure
500 ms
200 ms
500 ms
200 ms
TIME
96 blocks (42 trials in each, altogether 4032 trials)
12 epochs
D
Between-groups design
Structured-first group
48 structured blocks
48 unstructured blocks
Unstructured-first group
48 unstructured blocks
48 structured blocks


A
Structured-first group: Structured epochs
High
Low
RT (ms)
430
410
390
370
350
1
2
3
4
5
6
Epochs
B
Structured-first group: Unstructured epochs
RT (ms)
430
410
390
370
350
1
2
3
4
5
6
Epochs
C
Unstructured-first group: Unstructured epochs
RT (ms)
430
410
390
370
350
1
2
3
4
5
6
Epochs
D
Unstructured-first group: Structured epochs
RT (ms)
430
410
390
370
350
1
2
3
4
5
6
Epochs


Structured-first group
Unstructured-first group
p < .001
p < .001
Triplet learning score (ms)
10
8
6
4
2
0
-2
Structured epochs
Unstructured epochs