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 [d,c,f,*,+], Karolina Janacsek [d,c,g,*]

[a]: Brain Imaging Centre, Research Centre for Natural 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, *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), Université de Lyon, *Centre Hospitalier Le Vinatier – Bâtiment 462 – Neurocampus 95 Boulevard Pinel 69675 Bron, Lyon, France*

[g]: School of Human Sciences, Faculty of Education, Health and Human Sciences, University of Greenwich, *Old Royal Naval College, Park Row, 150 Dreadnought, London, United Kingdom*

*: These authors contributed equally to this work.

+: Corresponding authors:
Andrea Kóbor
Brain Imaging Centre, Research Centre for Natural 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 unclear to what extent implicit prior knowledge is involved and remains persistent 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, more probable and less probable short-range nonadjacent transitional probabilities were present. In the unstructured blocks, 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 half of the participants completed the structured blocks first followed by the unstructured blocks, this was reversed in the other half of them. The change of the structure was not explicitly denoted, and no feedback was provided upon the correctness of each response. Participants completing the structured blocks first showed faster RTs to more probable than to less probable short-range transitional probabilities in both the structured and unstructured blocks, indicating the persistent effect of prior knowledge. In contrast, participants completing the unstructured blocks first showed this RT effect only in the structured blocks. These results suggest that the implicitly acquired prior knowledge on a transitional probability structure influences the processing of unstructured stimuli through the formation of robust internal models. The effect of these internal models on processing remains relatively persistent even after they lost their relevance, suggesting the fundamental role of transitional probabilities in human cognition.

Keywords: implicit statistical learning, internal models, prior knowledge, transitional probabilities

1. 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 using previous experiences. However, the potential stability of the already acquired implicit knowledge when applied in similar situations has not been completely elucidated (Bulgarelli & Weiss, 2016; Conway, 2020; R. Frost, Armstrong, & Christiansen, 2019). Therefore, we investigate the stability of implicit knowledge on a statistical structure underlying a stream of visual stimuli that remains the same at the surface level but, in time, becomes unpredictable at the structural level.

According to the broad frameworks of cognitive processing, learning, and decision making, the processing of new information and the forming of expectations about future events are guided by inferences based on prior experiences (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 (Conway, 2020). For instance, the experienced features of statistical structures spontaneously occurring in random sequences influence randomness perception and random sequence generation (Hahn & Warren, 2009; Sun et al., 2015; Sun & Wang, 2010; Teigen & Keren, 2020; Warren, Gostoli, Farmer, El-Deredy, & Hahn, 2018).

Findings of binary choice experiments also underpin 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) has been linked to the tendency to search for an underlying statistical structure (e.g., Feher da Silva & Baldo, 2012; Gaissmaier & Schooler, 2008). It has also been observed that the generation of sequence-wide expectations increased the frequency of this behavior (James & Koehler, 2011). The persistency of the primarily acquired statistical structure and its influence on further processing have been evidenced by behavioral (e.g., Bulgarelli & Weiss, 2016; Gebhart, Aslin, & Newport, 2009; Lany, Gómez, & Gerken, 2007) as well as neurocognitive measures (e.g., Honbolygó & Csépe, 2013; Karuza et al., 2016; Mullens et al., 2014; Todd, Frost, Fitzgerald, & Winkler, 2020; Todd, Provost, & Cooper, 2011). Overall, previous findings highlight the role of prior internal models in the implicit processing of statistical structures (Qian, Jaeger, & Aslin, 2012) that can be different in characteristics and complexity (Conway, 2020).

Different statistical structures can be acquired from the same stimulus sequence (Conway & Christiansen, 2001; Daltrozzo & Conway, 2014; R. Frost et al., 2019). 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 Maheu et al., 2019; Orbán, Fiser, Aslin, & Lengyel, 2008). Humans have been found to be highly proficient in extracting even the nonadjacent transitional probabilities referring to predictive relations between elements of a sequence that includes ordered stimuli interspersed with random ones (Conway, 2020; R. L. A. Frost & Monaghan, 2016; Malassis, Rey, & Fagot, 2018; Mueller, Milne, & Männel, 2018; Rey, Minier, Malassis, Bogaerts, & Fagot, 2018).

Using transitional probabilities in a series of experiments, Gebhart et al. (2009) changed the underlying statistical structure of stimuli in the middle of an auditory statistical learning task. They successively presented two different but overlapping artificial speech streams composed of trisyllabic nonsense words characterized by transitional probabilities. This way, the surface of the stimuli remained similar throughout the task while their structure changed. If the change was not explicitly signaled or the second structure was not presented for a tripled duration, participants only learned the first structure. This indicated that the primarily experienced structure limited the capacity to acquire the successive structure, supporting the role of prior internal models. However, in this study, a certain statistical structure determined by transitional probabilities was always present during the task (see also Bulgarelli & Weiss, 2016; Weiss, Gerfen, & Mitchel, 2009; Zinszer & Weiss, 2013). Therefore, it is unclear whether the results would have been the same if the statistical structure per se had been eliminated. For instance, it could be clarified whether changing only the predictability of the same nonadjacent transitional probabilities over the course of learning influences their later extraction.

Furthermore, in the Gebhart et al. (2009) study, after exposure to the speech stream, knowledge on the statistical structures was measured with two-alternative forced choice test trials in which familiarity judgements were provided. Meanwhile, processing-based or "online" tasks should be favored, since these tasks more likely reflect implicitly acquired statistical knowledge about which no consciously accessible representations are available. These tasks could also capture the trajectory of acquisition and provide information about the stability of the underlying processes when these processes actually operate (Christiansen, 2018; R. Frost et al., 2019). Therefore, it remains to be tested with an online, unsupervised statistical learning task (Fiser & Aslin, 2001; Qian et al., 2012) how changing the predictability of nonadjacent transitional probabilities impacts further acquisition.

Consequently, in the present study, we used a specific four-choice reaction time (RT) task to online measure the implicit processing and acquisition of a sequence composed of 2nd order nonadjacent transitional probabilities. In this sequence underlying the RT task, elements in position $n - 2$ predicted elements in position n with high or low probability. Unknown to participants, half of the task blocks included an alternating regularity between nonadjacent 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 appeared with equal probability. The task blocks were labeled as structured and unstructured blocks, according to the presence or absence of the alternating regularity. While one half of fifty participants completed the structured blocks first followed by the unstructured blocks, the other half of the participants completed the unstructured blocks first followed by the structured blocks. Participants received neither explicit information on the midstream change in structure nor feedback upon the correctness of each response throughout the task.

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 (cf. Zhao et al., 2019).

2. Material and methods

2.1 Participants

Fifty healthy young adults took part in the experiment.[1] 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 were 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.

[1] ASRT-studies with effect size measures for the prior knowledge effect were not available. Therefore, when determining the sample size per group, we followed the guidelines set by some of the previous behavioral ASRT-studies (Hallgató, Győri-Dani, Pekár, Janacsek, & Nemeth, 2013; Horváth, Török, Pesthy, Nemeth, & Janacsek, 2019; Nemeth, Hallgato, Janacsek, Sandor, & Londe, 2009; Nemeth & Janacsek, 2011; Nemeth, Janacsek, & Fiser, 2013; Szegedi-Hallgató et al., 2017; Vékony et al., 2019). In average, the sample size in these studies was approx. 23 per group (SD = 10.6).

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.

2.2 Experimental task

2.2.1 The Alternating Serial Reaction Time (ASRT) task

Implicit acquisition of 2nd order transitional probabilities was measured by a modified version of the ASRT task (J. H. Howard, Jr. & Howard, 1997; 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 corresponded 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 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 (J. H. Howard, Jr. & Howard, 1997). 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 (see also Figs. S1–S2). 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. This characteristic of the task ensures that sensitivity to a biased distribution of triplets can be quantified. 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 (see Figs. S1–S2).

2.2.2 Generation and selection of the unstructured sequences

Besides the *structured* ASRT sequences that included the alternating regularity, *unstructured* sequences were used, in which the alternating regularity was absent. The unstructured sequences had to meet two requirements. *First*, 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. Therefore, each of the 64 triplets had to occur 30 times in any of the unstructured sequences but without the presence and repetition of the alternating regularity (1920 triplets in total, see the second requirement). This way, unstructured sequences could also be considered as pseudorandom sequences with the constraint that all triplets occurred with equal probability (25%). *Second*, unstructured sequences had to contain the same number of trials as structured sequences because 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). 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 the 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). Indeed, the probability of triplets differed only in the structured sequences and their probability was equal in the unstructured sequences (see Figs. S1–S2). Meanwhile, in the remainder of the paper, triplets in the unstructured sequences are still referred to as either high- or low-probability according to their actual probability in the structured sequences.

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).

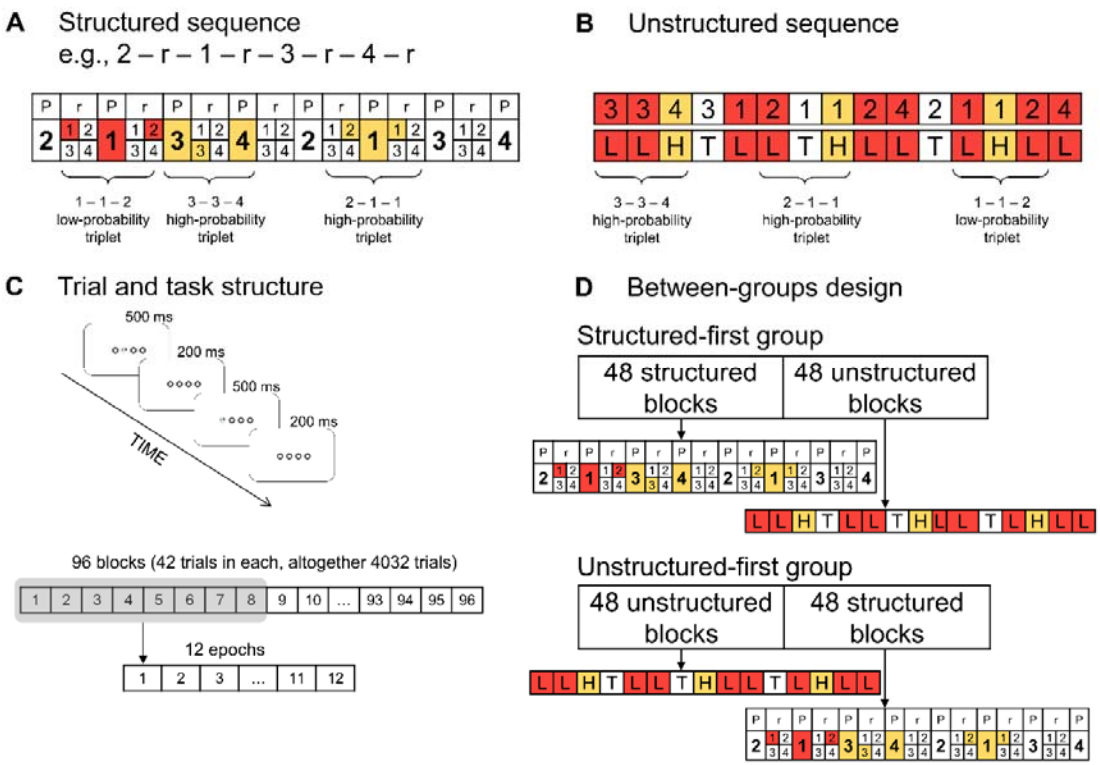

Figure 1. Design of the experiment. (A) The presentation of stimuli in the structured ASRT sequence followed an eight-element regularity, within which pattern (P) and random (r) elements alternated with one another. Numbers denote the four different stimulus positions on the screen. The alternating regularity 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. Note that the probability of triplets only differs in the structured sequence and their probability is equal in the unstructured sequence. However, triplets in the unstructured sequence are still labeled as either high- or low-probability according to their actual probability in the structured sequence. 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.

2.3 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 (keypress) 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 about their mean reaction time and accuracy in the given block. The length of this between-blocks "rest period" with feedback was jittered to be methodologically optimal for a future fMRI experiment (it lasted for 10, 12, or 14 secs [mean = 12 sec]). 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 version 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.

2.4 Statistical analysis

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 on the RTs, i.e., faster RTs to high-probability than to low-probability triplets, 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. Second, to more directly test the change in triplet learning as a function of the different 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. 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 pairwise comparisons were used to control for Type I error.

Regarding the possible experimental effects and their interpretation, the Triplet main effect implies *triplet learning* (faster RTs to high- than to low-probability triplets) and the Triplet * Epoch interaction implies *changes in triplet learning* as the task progresses, usually an increase across epochs (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). The Epoch main effect implies *general skill (RT) improvements* reflecting more efficient visuomotor and motor-motor coordination due to practice (Hallgató et al., 2013; Juhasz, Nemeth, & Janacsek, 2019). The *prior knowledge effect* is indicated by the

Sequence * Triplet * Group and/or the Sequence * Triplet * Epoch * Group interactions: Namely, if prior knowledge on the transitional probabilities influences later stimulus processing, triplet learning per se or its change over time should differ between structured and unstructured epochs *and* across the two groups experiencing the structured and unstructured epochs in an opposite order. In the Results section below, we use these terms when describing the observed statistical effects.

To follow up the prior knowledge effect, triplet learning scores in the structured and unstructured epochs were calculated as the RT difference between the 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. The overall triplet learning scores for the structured and unstructured epochs were first compared within each group. Then, these scores were compared between the groups. Finally, the change of mean RTs across the structured and unstructured epochs separately for the high- and low-probability triplets were compared within each group.

3. Results

3.1 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, $F(5, 240) = 35.91$, $\varepsilon = .612$, $p < .001$, $\eta_p^2 = .428$. As these main effects were qualified by significant higher-order interactions, only the latter effects are detailed below.

3.1.1 Triplet learning

The Sequence * Triplet, $F(1, 48) = 22.92$, $p < .001$, $\eta_p^2 = .323$, and the Triplet * Epoch, $F(5, 240) = 2.42$, $p = .036$, $\eta_p^2 = .048$, interactions were significant, while the Sequence * Triplet * Epoch, $F(5, 240) = 2.18$, $\varepsilon = .814$, $p = .072$, $\eta_p^2 = .043$ interaction was a trend level. These effects indicated that the change of triplet learning over the course of the task differed between the structured and unstructured epochs.

3.1.2 General skill improvements

The significant Sequence * Group, $F(1, 48) = 69.87$, $p < .001$, $\eta_p^2 = .593$, and the Sequence * Epoch * Group, $F(5, 240) = 33.59$, $\varepsilon = .706$, $p < .001$, $\eta_p^2 = .412$, interactions 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 this half being structured or unstructured and the distribution of triplets. Relatedly, the following *nonsignificant* main effects and interactions emerged: main effect of Group, $F(1, 48) = 1.96$, $p = .168$, $\eta_p^2 = .039$, Epoch * Group interaction, $F(5, 240) = 0.47$, $\varepsilon = .612$, $p = .706$, $\eta_p^2 = .010$, and Sequence * Epoch interaction, $F(5, 240) = 1.20$, $\varepsilon = .706$, $p = .313$, $\eta_p^2 = .024$.

3.1.3 Prior knowledge effect

The significant Triplet * Group interaction, $F(1, 48) = 4.92$, $p = .031$, $\eta_p^2 = .093$, was qualified by the significant Sequence * Triplet * Group interaction, $F(1, 48) = 7.96$, $p = .007$, $\eta_p^2 = .142$. Importantly, the latter indicated that the difference in triplet learning between the

structured and unstructured epochs varied across the groups, which is regarded as the prior knowledge effect (see Fig. 2). This effect did not reliably vary as a function of practice with the task as shown by the nonsignificant Sequence * Triplet * Epoch * Group interaction, $F(5, 240) = 0.61$, $\varepsilon = .814$, $p = .661$, $\eta_p^2 = .012$. Relatedly, the modulating effect of structured vs. unstructured epochs on triplet learning was also supported by the significant Triplet * Epoch * Group interaction, $F(5, 240) = 2.31$, $p = .045$, $\eta_p^2 = .046$, showing that if both task halves with the structured and unstructured epochs were collapsed, the trajectory of triplet learning would differ across the groups.

3.2 Follow-up of the prior knowledge effect

To follow up the prior knowledge effect, pairwise comparisons contrasting the overall triplet learning scores were performed (see Fig. 3). In the Structured-first group, the triplet learning score was similar between the structured and unstructured epochs (6.3 ms vs. 4.6 ms, $p = .171$). Thus, the behavioral effect of biased triplet probabilities (i.e., high- and low-probability triplets in the structured epochs) persisted even after this bias was eliminated (i.e., the unique triplets occurred with equal probability in the unstructured epochs). Meanwhile, in the Unstructured-first group, the triplet learning score was significantly larger over the structured epochs than over the unstructured epochs (5.9 ms vs. -0.4 ms, $p < .001$). In addition, the triplet learning score 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], $p < .001$; the difference in RT decrease between high- and low-probability

triplets was not significant, $p = .226$), indicating only general skill improvements from the structured to the unstructured epochs. In contrast, 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], $p < .001$; the difference in RT decrease between high- and low-probability triplets was significant, $p < .001$) from the unstructured to the structured epochs, indicating triplet learning over the structured epochs.

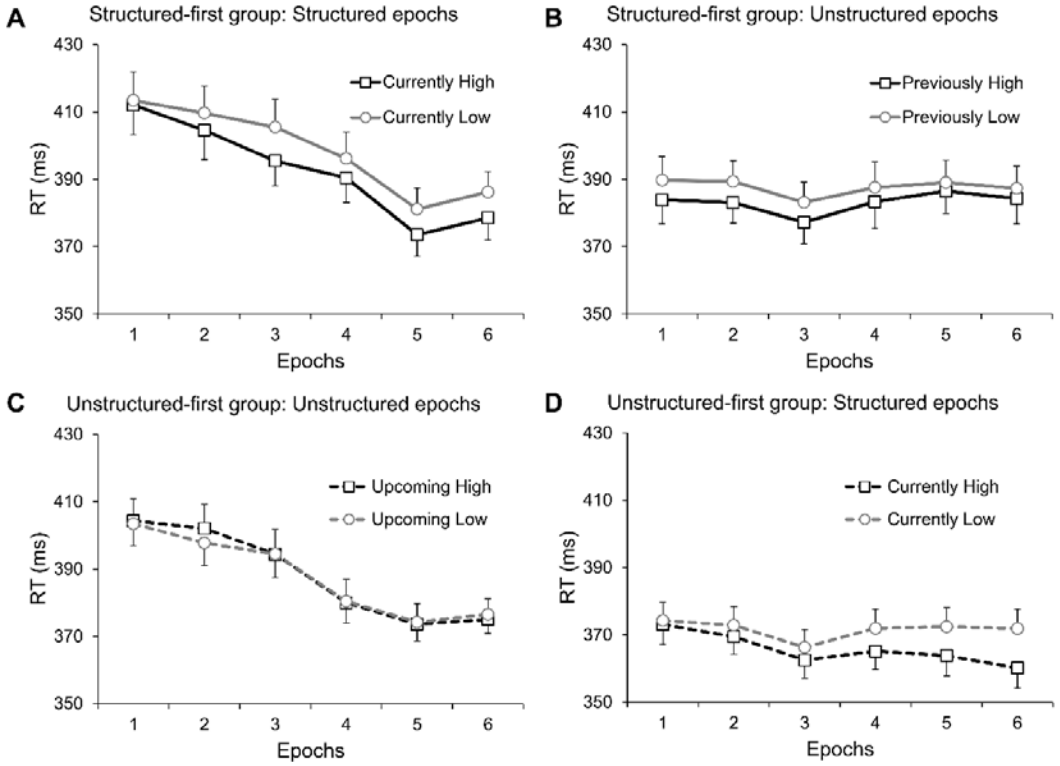

Figure 2. Temporal dynamics of triplet learning across groups and sequence types. Group-average RTs (A–B: Structured-first group; C–D: Unstructured-first group) for correct responses as a function of time bin (epochs 1–6) and triplet type (currently/previously/upcoming high- vs. low-probability triplets,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

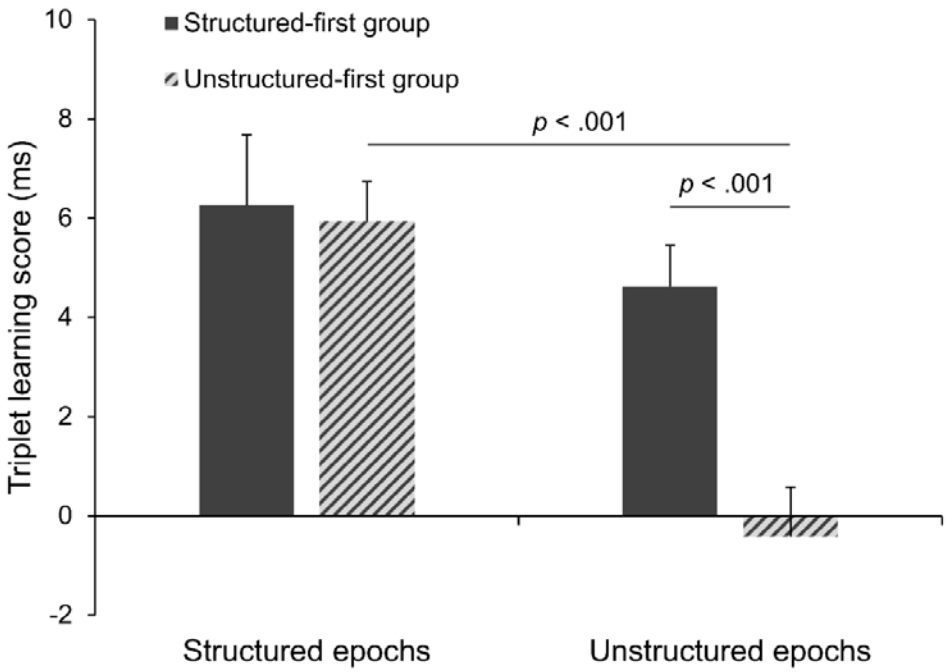

Figure 3. Persistence of the acquired implicit knowledge. Group-average overall (mean) triplet learning scores (RTs to low- minus RTs to high-probability triplets) are presented in the Structured-first and Unstructured-first groups for the structured and unstructured epochs, respectively. Error bars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

3.3 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). Triplet learning and its change over the structured epochs did not differ between the groups as shown by the nonsignificant Triplet * Group, $F(1,$

48) = 0.04, p = .840, η_p^2 = .001, and Triplet * Epoch * Group interactions, F(5, 240) = 1.54, p = .177, η_p^2 = .031.

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, η_p^2 = .181, and Epoch, F(5, 240) = 13.45, ε = .605, p < .001, η_p^2 = .219, while the Triplet * Epoch interaction was not significant, F(5, 240) = 0.24, p = .945, η_p^2 = .005). Importantly, the significant Triplet * Group interaction, F(1, 48) = 15.23, p < .001, η_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, but this did not change with time (nonsignificant Triplet * Epoch * Group interaction, F(5, 240) = 1.33, p = .251, η_p^2 = .027). The trajectory of general skill improvements differed between the groups (significant Epoch * Group interaction, F(5, 240) = 16.12, ε = .605, p < .001, η_p^2 = .251, see Fig. 2B, C), and the groups did not differ in overall RT (nonsignificant Group main effect, F(1, 48) = 0.09, p = .764, η_p^2 = .002) over the unstructured epochs.

3.4 Analysis of accuracy

Since only the RTs of the correctly responded trials were analyzed, it should be ensured that the two groups did not differ in accuracy. Therefore, the Sequence by Triplet by Epoch by Group ANOVA was conducted also on accuracy data (calculated as the ratio of correct responses for each participant and epoch, separately for high- and low-probability triplets). As indicated by the nonsignificant main effect of Group, F(1, 48) = 1.53, p = .221, η_p^2 = .031, the two groups were comparable on overall accuracy (Structured-first group: 86.4%; Unstructured-first group: 88.1%).

Although the analysis of accuracy is not the focus of this study, for the sake of completeness, we provide the other significant main effects and interactions revealed in this

ANOVA, but these are not detailed. The main effects of Triplet, $F(1, 48) = 25.25$, $p < .001$, $\eta_p^2 = .345$, and Epoch, $F(5, 240) = 7.42$, $\varepsilon = .639$, $p < .001$, $\eta_p^2 = .134$, were significant. Relatedly, the Sequence * Triplet, $F(1, 48) = 4.91$, $p = .032$, $\eta_p^2 = .093$, and the Sequence * Triplet * Epoch, $F(5, 240) = 3.49$, $p = .005$, $\eta_p^2 = .068$, interactions were also significant. In brief, responses to high-probability triplets were more accurate than to low-probability ones. However, while this difference in accuracy increased over the structured epochs, it gradually decreased over the unstructured epochs.

The Sequence * Group, $F(1, 48) = 9.75$, $p = .003$, $\eta_p^2 = .169$, and the Sequence * Epoch * Group, $F(5, 240) = 7.44$, $\varepsilon = .524$, $p < .001$, $\eta_p^2 = .134$, interactions were significant, as well. The Sequence * Triplet * Epoch * Group interaction was a trend level, $F(5, 240) = 1.92$, $p = .091$, $\eta_p^2 = .039$. The latter effect indicated that the above-described Sequence * Triplet * Epoch interaction was mostly driven by the responding pattern of the Structured-first group. Particularly, in the Structured-first group, while the difference in accuracy between high- and low-probability triplets increased over the structured epochs, it tended to decrease over the unstructured epochs. In the Unstructured-first group, accuracy between high- and low-probability triplets only differed in the structured epochs.

4. Discussion

4.1 Summary of results

This study investigated whether the implicitly acquired knowledge on a 2nd order transitional probability structure influenced the processing of unpredictable transitional probabilities across phases of a learning task. 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 biased triplet

probabilities were 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 already acquired implicit knowledge on the short-range transitional probabilities has been persistent across the learning phases. Importantly, this prior knowledge influenced the processing of a stimulus stream with an equal transitional probability structure, consisting only of uniformly distributed stimulus chunks that otherwise could not be extracted.

4.2 Interpretation of results: Robust implicit statistical knowledge

Participants completing the structured blocks first similarly perceived the relations of stimuli in both type of blocks. That is, their perception could have been influenced by the primarily experienced transitional probability structure. In other words, because of the task environment, they might have worked up a tendency towards pattern detection, which could have resulted in forming implicit expectations about the upcoming stimuli. Then, these expectations remained persistent throughout the task.

This persistency is in line with findings from a binary choice task testing probability matching, where, in contrast to the present study, the transition between the no pattern (no serial dependence in the sequence) and pattern half (repeating deterministic sequence) was clearly indicated. Results showed 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 (Gaissmaier &

Schooler, 2008). Meanwhile, it was also found previously that individuals accurately estimated the hidden probability parameter of a nonstationary stochastic environment as well as quickly updated their estimate (Gallistel, Krishan, Liu, Miller, & Latham, 2014). In this task, participants assessed the proportion of one stimulus category and had the opportunity to update their estimate on a trial-by-trial basis. Importantly, at the beginning of the task, they were told that probabilities could unexpectedly change. Similarly, in another experiment where subsequent numerical values had to be predicted, participants updated each prediction as a function of their explicitly denoted prediction error. By tracking the prediction errors, after an unsigned change in the distribution of the values during the task, participants could adjust their predictions (Nassar, Wilson, Heasly, & Gold, 2010). The quick adaptation to changing probabilities was also observed when choosing between two options associated with different probabilities and reward magnitudes, i.e., with clear feedback signals (Behrens, Woolrich, Walton, & Rushworth, 2007). These studies suggest that effective decision making necessitates the continuous tracking of the environmental probabilities and the evaluation of each signal that possibly implies a change in these probabilities. However, these observations have been derived from situations in which active agents were required to make explicit decisions that pertained directly to the probabilistic features of the ongoing task modeling volatile environments.

In contrast, in the present ASRT task, acquisition happened in an incidental and implicit manner: our participants did not know that they were in a learning situation, they did not have information on whether the sequence of stimuli was random or followed any underlying pattern, the critical change point between the task halves remained unnoticed, and they were not required to actively or explicitly predict the probability of the next stimulus. In addition, they did not receive feedback (or any reward) on the correctness of their responses. Meanwhile, participants were instructed to maintain a certain level of speed and accuracy,

but, for them, this was the only explicit goal of the task. They successfully achieved this goal, as shown by the behavioral results indicating general skill improvements due to practice. This performance improvement was mostly independent of the change in the underlying probability structure. Thus, it seems that as participants had gained confidence in task solving, they followed their already established, automatized strategy on stimulus processing and responding (cf. Karuza et al., 2016). Therefore, in our interpretation, as the surface of the task remained consistent, participants had no reason to doubt their current implicit beliefs about the probability of the upcoming stimulus (cf. Zinszer & Weiss, 2013). (Also, some of the participants in both groups might have inherently possessed a baseline tendency towards a continuous pattern detection, but this was not investigated here.)

Because of these characteristics, the present paradigm should not be considered as a(n) (explicit) decision-making or probabilistic reinforcement-learning paradigm in which quick updating of beliefs could happen (Bulgarelli & Weiss, 2016). Instead, the type of learning that this task measures more likely fits into the category of unsupervised statistical learning (Fiser & Aslin, 2001). It intends to model a stable stimulus environment with low volatility where hidden probabilistic regularities occur interspersed with noise. In addition, the implicit acquisition of more probable and less probable nonadjacent transitional probabilities also differentiates the ASRT task from the above-mentioned paradigms. The extraction of long-distance dependencies such as the nonadjacent transitional probabilities is assumed to be a relatively complex, predictive problem supported by various neurocognitive processes (Bennett, Madden, Vaidya, Howard, & Howard, 2011; Conway, 2020; R. Frost et al., 2019; Karuza et al., 2016; Kóbor et al., 2019; Malassis et al., 2018; Stillman et al., 2013; Tóth et al., 2017).

Such complex representations characterize, for instance, the native or current language environment and those automatic movements, actions, and perceptual-cognitive operations

that are related to usual, day-to-day activities (Conway, 2020; R. Frost et al., 2019; Meyniel et al., 2016; Saffran, Aslin, & Newport, 1996). Regarding the widespread occurrence and complexity of nonadjacent transitional probabilities, once their extraction has taken place, it could be adaptive that the related cognitive/perceptual representations remain persistent over time and robust to change. This way, the representations could remain sensitive to the primary transitional probability structure later in time, albeit other structures are simultaneously acquired (cf. Gebhart et al., 2009; Qian et al., 2012; Todd et al., 2020; Todd et al., 2011). This might be even more pronounced if no explicit cue or performance deterioration signals the need for a possibly effortful updating process. In our view, such a "processing efficiency" might explain our results on the transition from structured to pseudorandom stimuli; however, this is only an assumption and should be formally tested in future studies.

Supporting this notion, during the acquisition of different statistical structures, a tendency towards neural efficiency coupled with diminished sampling of the input underlying the second statistical structure has been shown (Karuza et al., 2016). Meanwhile, studies have also started to investigate how to attenuate the persistent effect of the already acquired statistical representations. By presenting two artificial speech streams in smaller alternating blocks, individuals were able to learn both statistical structures underlying the input streams, without using explicit contextual cues (e.g., change in speaker) denoting the transitions across the streams (Zinszer & Weiss, 2013). However, if statistically incongruent, interfering statistical structures determined the stimuli, the formation of multiple representations were limited (Weiss et al., 2009). Importantly, if individuals were exposed to the second statistical structure immediately after learning had occurred on the first structure presented for a restricted time, both structures were learned. In addition, the different contextual cues did not further enhance performance (Bulgarelli & Weiss, 2016). Overall, it seems that overlearning the first statistical structure and low variability in how the statistical structures are presented

could decrease the attention paid to the input stream, thereby deteriorating the acquisition of the new structure (Bulgarelli & Weiss, 2016). This explanation might be feasible in the case of our findings; however, it should be noted that these previous studies tested the transition(s) between different statistical structures in the linguistic domain, while our study investigated the unsigned transition from the presence to the absence of a structure in the visuomotor domain. Nevertheless, we can contribute to this research field by confirming the presence of the primacy effect in a unique multi-context unsupervised learning environment and, this way, by extending the validity of this effect to unlearnable pseudorandom stimuli.

In our study, the probabilistic nature of the applied statistical structure might have directly contributed to the findings on persistent implicit knowledge. Particularly, when a repeating deterministic sequence guides stimulus presentation in the structured blocks, such as in the case of the classical serial reaction time (SRT) task (e.g., Nissen & Bullemer, 1987), performance usually deteriorates on the unstructured blocks with random or pseudorandom stimuli. This means that RTs suddenly increase compared to the level reached by the end of the last structured block. Meanwhile, in the present study, participants completing the structured blocks first showed persistent triplet learning performance over several blocks of pseudorandom stimuli in terms of RTs. It is possible that in the case of probabilistic sequences as opposed to deterministic ones, the acquisition processes are more sensitive to smooth transitions between stimuli or chunks of stimuli. This specific sensitivity evolved in participants practicing the structured blocks first might have led the extraction of triplets even over the unstructured blocks of the task, where the triplets appeared with equal probability (see also the Transfer of prior knowledge section). In addition, as opposed to deterministic and pseudorandom sequences, the probabilistic sequence might have provided learnable but sufficiently novel information on a trial-by-trial basis, promoting the relatively fast acquisition of the statistical structure as well as its persistence.

4.3 The temporal scale of persistency

It has already been demonstrated that the primarily acquired implicit knowledge on the biased distribution of transitional probabilities 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 (that partially overlap with the primarily practiced sequence) not only after 24 hours but also after one year (Kóbor et al., 2017). We add to these results by providing evidence for the formed probabilistic representations being robust to an essential change in the probability model characterizing the given environment, at least over a short time period. However, based on the previous literature on the transitions between multiple statistical structures (Bulgarelli & Weiss, 2016; Gebhart et al., 2009; Karuza et al., 2016; Weiss et al., 2009; Weiss, Schwob, & Lebkuecher, 2019; Zinszer & Weiss, 2013), exact assumptions cannot be formed on the temporal dynamics of updating this existing implicit knowledge. These studies used linguistic stimuli with transitional probabilities differing in structure and complexity from that of the 2nd order nonadjacent transitional probabilities applied in the present task. In addition, the overall learning sessions for each statistical structure were much shorter (ranging from two-three minutes to 15 minutes) than in the present task (each sequence type were practiced for approx. 34 minutes). Accordingly, in our completely different experimental setting, it would be difficult to predict for how long the already acquired implicit knowledge would persist after the change in structure. Nevertheless, our results also highlight the constraining effect of the primarily acquired, possibly overlearned statistical structure in the adaptation to a new environment, as described below (cf. Bulgarelli & Weiss, 2016; Gebhart et al., 2009).

It has been suggested that successful adaptation to a range of similar tasks requires the forgetting or weakening of some specific features of the already acquired representations

(Robertson, 2018). In the present case, this would have been the forgetting of the initial triplet frequency information when starting the other task half. Particularly, it is probable that participants completing the structured blocks first 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, $t(24) = 2.46$, $p = .022$). In line with this, accuracy data for the previously high- and low-probability triplets are also approaching one another in the second half of the unstructured epochs, and a similar significant decrease in triplet knowledge measured by accuracy can be observed (2.7% vs. 0.7%, $t(24) = 2.91$, $p = .008$). These additional results might indicate an "implicit need" for updating the prior model on the probability structure. It is plausible to assume that after having had performed even more unstructured blocks, participants would have learned that the previously high-probability triplets no longer occurred with higher frequency than the previously low-probability ones, and, therefore, the initial triplet frequency information would have been forgotten or "unlearned".

However, 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, it is more likely that, at the level of triplet representations, no "extinction" or "unlearning" might have happened. Evidence from human and animal studies suggests that adaptation to different contexts does not involve the complete removal of representations on the previous context (Bulgarelli & Weiss, 2016; Gordon, Bilolikar, Hodhod, & Thomas, 2020; Qian et al., 2012). Instead, the forming of new representations, the reconsolidation or inhibition of the previously created ones, and the

switching between multiple representations seem more likely (Chandler & Gass, 2013). Either of the latter processes 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). The exact mechanisms by which this slow change might have occurred has yet to be determined and formal models should be developed to investigate the temporal dynamics of these mechanisms (cf. R. Frost et al., 2019; Karuza et al., 2016; Qian et al., 2012; Zhao et al., 2019).

With the help of the between-subjects design, we could provide further insights into the processes of implicitly acquiring a complex transitional probability structure and updating this knowledge. Namely, over the structured blocks, participants completing the unstructured blocks first showed a triplet learning trajectory comparable to that of the other group (nonsignificant Triplet * Group and Triplet * Epoch * Group interactions, see also Fig. 2A, D). With an unsigned change in the probability structure, the pervasive experience with pseudorandom stimuli could have also caused entrenchment effect: In this case, these participants could have started the acquisition of the biased probability structure with some disadvantage or could have showed the complete lack of triplet learning when exposed to the structured blocks. Instead, according to the results, it seems that prior experience with equal transitional probabilities did not influence negatively the further acquisition of the biased probability structure.

As an explanation for the triplet learning performance of these participants, it is conceivable that they might have primarily established a wider hypothesis space about the properties of the stimuli, since they could not extract complex transitional probabilities over the unstructured blocks. Such implicit internal models could be useful if the stimuli considered as random in the given environment with limited observations in fact followed

some structure (cf. Gaissmaier & Schooler, 2008). With a wider hypothesis space, stimulus processing and acquisition might have proceeded flexibly, enabling the acquisition of the statistical structure when it was indeed present. In addition, this triplet learning trajectory resembled the trajectory of updating over the unstructured blocks in the other group of participants (see Fig. 2B, D), which has not been previously suggested and might merit further investigation.

4.4 Transfer of prior knowledge

Considering the underlying processes, the present findings raise the question whether learning transfer has occurred across the task halves. Studies testing the transfer (generalization) of perceptual and motor knowledge usually compare performance observed during a training task with performance observed during a similar testing condition, such as in a familiar task with new parameters or in a related but novel task. Successful transfer occurs if experience gathered on the training task appears as a performance gain on the novel task (e.g., Dorfberger, Adi-Japha, & Karni, 2012; Korman, Raz, Flash, & Karni, 2003). For instance, a trained movement sequence can be transferred to the untrained hand or to a new, untrained sequence with the same hand; similarly, performance in a visual search task can be transferred across different array sizes (for review, see Karni, 1996; Karni & Bertini, 1997).

In the statistical-sequence learning literature, the implicit transfer of both the perceptual and the motor sequence were shown in a version of the ASRT task that, in the testing phase, included a novel, previously unpracticed alternating motor or perceptual sequence with the same type of stimuli (Hallgató et al., 2013; Nemeth et al., 2009). In a deterministic SRT task, the 2nd order transitional probability structure was implicitly transferred from the training to the testing phase where the perceptual features of the stimuli

differed (i.e., 1st order structure: locations arranged horizontally or according to a square, Huang et al., 2017).

Likewise, experiments using the artificial grammar learning (AGL) task found the implicit transfer of sequential dependencies to novel vocabularies (Tunney & Altmann, 2001). Other studies in this research field investigated the nature of representations based on which the transfer of artificial grammars occurs (e.g., Dienes & Altmann, 1997; Gomez, Gerken, & Schvaneveldt, 2000; Tunney & Altmann, 2001; Tunney & Shanks, 2003). Relatedly, unconscious within- and between-modalities transfer of artificial grammars was more recently shown between training and test strings that changed at the surface level (letters in different vocabularies, notes, symbols) but remained structurally the same (Scott & Dienes, 2010). Another line of research found learning transfer in both directions between different types of memory tasks (motor skill task and word list task) via the extraction of the high-level relations between the elements (Mosha & Robertson, 2016).

In contrast to these studies, the present experiment followed a different design: The surface of the stimuli remained the same across the two task phases, but their overall underlying structure changed. In this sense, the observed effect might not be considered as a classical perceptual-motor transfer effect. Indeed, it more likely captures a cognitive or model-based transfer effect: After short experience with the unstructured sequence, participants completing first 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., Qian et al., 2012; Robertson, 2018; Winkler & Cowan, 2005). Thus, the triplets as the critical building blocks of the structure might have been implicitly recalled in a later phase of task solving, even if their frequencies had changed. This would be in line with our recent findings that sensitivity to multiple regularities in the ASRT task seems to be grounded in the implicit extraction of the

triplet-level probability structure (e.g., Kóbor et al., 2019; Szegedi-Hallgató, Janacsek, & Nemeth, 2019). Together with the previously described various transfer effects, the implicitly acquired prior knowledge seems to be robust to changes in both the surface and the underlying structure of the stimuli.

4.5 An alternative approach to test the transfer of transitional probabilities

When primary experiences are gathered from a uniform distribution of stimuli, complex transitional probabilities could not be extracted, as implied by the results of the group completing the unstructured sequence first. Nevertheless, we cannot exclude the possibility that even without the alternating regularity, triplet learning would occur if triplet probabilities were changed otherwise. Considering this issue, an alternative experimental design with biased triplet probabilities would unequivocally test the transfer of the acquired implicit knowledge.

For instance, one might use the ASRT sequence as the “structured sequence” and create another “less structured” sequence by keeping the biased distribution of high- (62.5%) and low-probability (37.5%) triplets but omitting the alternating regularity. In a similar between-subjects design, by presenting either the structured or the less structured sequence in the first half of the task and the other sequence in the second half, a future study might compare triplet learning between these two sequences. To follow previous experiments testing the transfer effect, the stimuli in the second half of the task might be different at the surface level (e.g., arrows or colors instead of horizontally arranged positions), but this is not necessary (cf. Gebhart et al., 2009). In another version of this experiment, the less structured sequence could be presented in both task halves for one of the groups, while the structured sequence followed by the less structured sequence could be presented for the other group. In this version, stimuli should differ at the surface level in the second half of the task.

If triplet learning occurs in both sequences but only after participants completed the structured sequence first followed by the less structured one, it will indicate that knowledge on the transitional probability structure has been transferred across the task halves. However, for participants completing the less structured sequence first or completing only less structured sequences in both halves, we assume that a modest degree of triplet learning would occur also on these less structured sequences, because of the initial sensitivity to the triplet-level probability structure (see above). Clearly, this needs to be tested in additional experiments, because it is not confirmed that triplet learning is a general effect or specific to the ASRT task.

4.6 Methodological considerations

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. 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 testing sequence, one should use training and testing 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 both 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 first to the second sequence.

Therefore, in the present experiment, 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, Bogaerts, Christiansen, & Frost, 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 completing these blocks first, 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 support of this idea, other experiments indicated that low-probability events were considered as more random than high-probability events (Teigen & Keren, 2020), which might also pertain to the low-probability triplets that were the major constituents of the unstructured sequence in this study.

4.7 Conclusions

The present experiment provides evidence that under implicit and incidental learning conditions, humans continuously believe in prior internal models on a complex transitional probability structure, although, after a while, the predictable transitional probability structure does not exist. These findings could contribute to the understanding of how implicit prior knowledge influences further perceptual and cognitive processing. Particularly, they indicate that due to the robustness and persistency of the formed internal models, a stimulus environment lacking a predictable transitional probability structure is automatically processed in a model-based manner. Moreover, the findings also indicate that stable representations related to nonadjacent transitional probabilities occupy a crucial role in forming long-term

implicit memory traces. Finally, this study also highlights the importance of carefully constructing the underlying structure of training and testing 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 of the University of Lyon as part of the Programme Investissements d'Avenir (ANR-16-IDEX-0005 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, Borbála German for helping in data acquisition, Zsófia Zavecz for providing helpful comments on the experimental design and on an earlier version of this paper, and the reviewers for their illuminating comments and suggestions.

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

Behrens, T. E. J., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. S. (2007). Learning the value of information in an uncertain world. *Nature Neuroscience, 10*(9), 1214-1221. doi:10.1038/nn1954

Bennett, I. J., Madden, D. J., Vaidya, C. J., Howard, J. H., Jr., & Howard, D. V. (2011). White matter integrity correlates of implicit sequence learning in healthy aging. *Neurobiology of Aging, 32*(12), 2317 e2311-2312. doi:10.1016/j.neurobiolaging.2010.03.017

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

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

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. (2020). How does the brain learn environmental structure? Ten core principles for understanding the neurocognitive mechanisms of statistical learning. *Neuroscience & Biobehavioral Reviews*. doi:10.1016/j.neubiorev.2020.01.032

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

Dienes, Z., & Altmann, G. T. M. (1997). Transfer of implicit knowledge across domains: How implicit and how abstract? In D. C. Berry (Ed.), *How implicit is implicit learning?* (pp. 107-123). Oxford: Oxford University Press.

Dorfberger, S., Adi-Japha, E., & Karni, A. (2012). Sequence specific motor performance gains after memory consolidation in children and adolescents. *PLoS ONE, 7*(1), e28673. doi:10.1371/journal.pone.0028673

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

Fiser, J., & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science, 12*(6), 499-504. doi:10.1111/1467-9280.00392

Fiser, J., & Aslin, R. N. (2002). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences of the United States of America, 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

Frost, R., Armstrong, B. C., & Christiansen, M. H. (2019). Statistical learning research: A critical review and possible new directions. *Psychological Bulletin, 145*(12), 1128-1153. doi:10.1037/bul0000210

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

Gallistel, C. R., Krishan, M., Liu, Y., Miller, R., & Latham, P. E. (2014). The perception of probability. *Psychological Review, 121*(1), 96-123. doi:10.1037/a0035232

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

Gomez, R. L., Gerken, L., & Schvaneveldt, R. W. (2000). The basis of transfer in artificial grammar learning. *Memory & Cognition, 28*(2), 253-263. doi:10.3758/bf03213804

Gordon, L. T., Bilolikar, V. K., Hodhod, T., & Thomas, A. K. (2020). How prior testing impacts misinformation processing: A dual-task approach. *Memory & Cognition, 48*(2), 314-324. doi:10.3758/s13421-019-00970-0

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

Horváth, K., Török, C., Pesthy, O., Nemeth, D., & Janacsek, K. (2019). Intention to learn differentially affects subprocesses of procedural learning and consolidation: Evidence from a probabilistic sequence learning task. *bioRxiv*, 433243. doi:10.1101/433243

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

Huang, J., Dai, H., Ye, J., Zhu, C., Li, Y., & Liu, D. (2017). Impact of response stimulus interval on transfer of non-local dependent rules in implicit learning: An ERP investigation. *Frontiers in Psychology, 8*(2107). doi:10.3389/fpsyg.2017.02107

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

Juhasz, D., Nemeth, D., & Janacsek, K. (2019). Is there more room to improve? The lifespan trajectory of procedural learning and its relationship to the between- and within-group differences in average response times. *PLoS ONE, 14*(7), e0215116. doi:10.1371/journal.pone.0215116

Karni, A. (1996). The acquisition of perceptual and motor skills: A memory system in the adult human cortex. *Cognitive Brain Research, 5*(1), 39-48. doi:10.1016/S0926-6410(96)00039-0

Karni, A., & Bertini, G. (1997). Learning perceptual skills: Behavioral probes into adult cortical plasticity. *Current Opinion in Neurobiology, 7*(4), 530-535. doi:10.1016/S0959-4388(97)80033-5

Karuza, E. A., Li, P., Weiss, D. J., Bulgarelli, F., Zinszer, B. D., & Aslin, R. N. (2016). Sampling over nonuniform distributions: A neural efficiency account of the primacy

effect in statistical learning. *Journal of Cognitive Neuroscience, 28*(10), 1484-1500. doi:10.1162/jocn_a_00990

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., Horváth, K., Kardos, Z., Takács, Á., Janacsek, K., Csépe, V., & Nemeth, D. (2019). Tracking the implicit acquisition of nonadjacent transitional probabilities by ERPs. *Memory & Cognition, 47*(8), 1546-1566. doi:10.3758/s13421-019-00949-x

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

Korman, M., Raz, N., Flash, T., & Karni, A. (2003). Multiple shifts in the representation of a motor sequence during the acquisition of skilled performance. *Proceedings of the National Academy of Sciences of the United States of America, 100*(21), 12492-12497. doi:10.1073/pnas.2035019100

Lany, J., Gómez, R. L., & Gerken, L. A. (2007). The role of prior experience in language acquisition. *Cognitive Science, 31*(3), 481-507. doi:10.1080/15326900701326584

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

Nassar, M. R., Wilson, R. C., Heasly, B., & Gold, J. I. (2010). An approximately bayesian delta-rule model explains the dynamics of belief updating in a changing environment. *The Journal of Neuroscience, 30*(37), 12366. doi:10.1523/JNEUROSCI.0822-10.2010

Nemeth, D., Hallgato, E., Janacsek, K., Sandor, T., & Londe, Z. (2009). Perceptual and motor factors of implicit skill learning. *Neuroreport, 20*(18), 1654-1658. doi:10.1097/WNR.0b013e328333ba08

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

Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology, 19*(1), 1-32. doi:10.1016/0010-0285(87)90002-8

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 of the United States of America, 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

Qian, T., Jaeger, T. F., & Aslin, R. (2012). Learning to represent a multi-context environment: More than detecting changes. *Frontiers in Psychology, 3*(228). doi:10.3389/fpsyg.2012.00228

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

Scott, R. B., & Dienes, Z. (2010). Knowledge applied to new domains: The unconscious succeeds where the conscious fails. *Consciousness and Cognition, 19*(1), 391-398. doi:10.1016/j.concog.2009.11.009

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

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

Stillman, C. M., Gordon, E. M., Simon, J. R., Vaidya, C. J., Howard, D. V., & Howard, J. H., Jr. (2013). Caudate resting connectivity predicts implicit probabilistic sequence learning. *Brain Connectivity, 3*(6), 601-610. doi:10.1089/brain.2013.0169

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., & Nemeth, D. (2019). Different levels of statistical learning - hidden potentials of sequence learning tasks. *PLoS ONE, 14*(9), e0221966. doi:10.1371/journal.pone.0221966

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

Teigen, K. H., & Keren, G. (2020). Are random events perceived as rare? On the relationship between perceived randomness and outcome probability. *Memory & Cognition, 48*(2), 299-313. doi:10.3758/s13421-019-01011-6

Todd, J., Frost, J., Fitzgerald, K., & Winkler, I. (2020). Setting precedent: Initial feature variability affects the subsequent precision of regularly varying sound contexts. *Psychophysiology, n/a*(n/a), e13528. doi:10.1111/psyp.13528

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

Tunney, R. J., & Altmann, G. T. (2001). Two modes of transfer in artificial grammar learning. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 27*(3), 614-639. doi:10.1037/0278-7393.27.3.614

Tunney, R. J., & Shanks, D. R. (2003). Subjective measures of awareness and implicit cognition. *Memory & Cognition, 31*(7), 1060-1071. doi:10.3758/bf03196127

Vékony, T., Marossy, H., Must, A., Vécsei, L., Janacsek, K., & Németh, D. (2019). Disentangling competence from performance in behavioral measures of learning: A lesson for cognitive neuroscience. *bioRxiv*, 726315. doi:10.1101/726315

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

Weiss, D. J., Gerfen, C., & Mitchel, A. D. (2009). Speech segmentation in a simulated bilingual environment: A challenge for statistical learning? *Language Learning and Development, 5*(1), 30-49. doi:10.1080/15475440802340101

Weiss, D. J., Schwob, N., & Lebkuecher, A. L. (2019). Bilingualism and statistical learning: Lessons from studies using artificial languages. *Bilingualism: Language and Cognition, 23*(1), 92-97. doi:10.1017/S1366728919000579

Winkler, I., & Cowan, N. (2005). From sensory to long-term memory: Evidence from auditory memory reactivation studies. *Experimental Psychology, 52*(1), 3-20. doi:10.1027/1618-3169.52.1.3

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.

Zhao, S., Chait, M., Dick, F., Dayan, P., Furukawa, S., & Liao, H.-I. (2019). Pupil-linked phasic arousal evoked by violation but not emergence of regularity within rapid sound sequences. *Nature Communications, 10*(1), 4030. doi:10.1038/s41467-019-12048-1

Zinszer, B., & Weiss, D. (2013). When to hold and when to fold: Detecting structural changes in statistical learning. *Proceedings of the Annual Meeting of the Cognitive Science Society, 35*, 3858-3863. Retrieved from https://escholarship.org/content/qt6rf3g020/qt6rf3g020.pdf