Evidence for a competitive relationship between executive functions and statistical learning

Felipe Pedraza[1,2*], Bence C. Farkas[3,4,5*], Teodóra Vékony[2+], Frederic Haesebaert[2], Romane Phelipon[2], Imola Mihalecz[2], Karolina Janacsek[6,7], Royce Anders[8], Barbara Tillmann[2,9], Gaën Plancher[1,10] & Dezso Nemeth[2,7,11+]

1 Laboratoire d'Étude des Mécanismes Cognitifs, Université Lumière Lyon 2, Bron, France

2 INSERM, Université Claude Bernard Lyon 1, CNRS, , Centre de Recherche en Neurosciences de Lyon CRNL U1028 UMR5292, 95 Boulevard Pinel, F-69500, Bron, France

3 Université Paris-Saclay, UVSQ, Inserm, CESP, 94807, Villejuif, France

4 Institut du Psychotraumatisme de l'Enfant et de l'Adolescent, Conseil Départemental Yvelines et Hauts-de-Seine et Centre Hospitalier des Versailles, 78000, Versailles, France

5 Centre de recherche en épidémiologie et en santé des populations, Inserm U1018, Université Paris-Saclay, Université Versailles Saint-Quentin, Paris, France

6 Centre for Thinking and Learning, Institute for Lifecourse Development, School of Human Sciences, Faculty of Education, Health and Human Sciences, University of Greenwich, Old Royal Naval College, Park Row, 150 Dreadnought, London, SE10 9LS, UK

7 Institute of Psychology, ELTE Eötvös Loránd University, Kazinczy u. 23-27, H-1075, Budapest, Hungary

8 EPSYLON Laboratory, Department of Psychology, University Paul Valéry Montpellier 3, F34000 Montpellier, France

9 Laboratory for Research on Learning and Development, LEAD – CNRS UMR5022, Université de Bourgogne, Dijon, France

10 Institut Universitaire de France (IUF), Paris, France

11 Brain, Memory and Language Research Group, Institute of Cognitive Neuroscience and Psychology, Research Centre for Natural Sciences, H-1117, Budapest, Hungary

*****These authors contributed equally to this work

[+]Corresponding authors. Address correspondence to Dezso Nemeth, Lyon Neuroscience Research Center (CRNL), INSERM, CNRS, Université de Lyon, Centre Hospitalier Le

Vinatier - Bâtiment 462 - Neurocampus 95 boulevard Pinel, 69675 Bron, France. E-mail: dezso.nemeth@univ-lyon1.fr or Teodóra Vékony, teodora.vekony@inserm.fr Lyon Neuroscience Research Center (CRNL), INSERM, CNRS, Université de Lyon, Centre Hospitalier Le Vinatier - Bâtiment 462 - Neurocampus 95 boulevard Pinel, 69675

The ability of the brain to extract patterns from the environment and predict future events, known as statistical learning, has been proposed to interact in a competitive manner with prefrontal lobe related networks and their characteristic cognitive or executive functions. However, it remains unclear whether these cognitive functions also show competitive relationship with implicit statistical learning across individuals and at the level of latent executive function components. In order to address this currently unknown aspect, we investigated, in two independent experiments (N_{Study1} = 186, N_{Study2} = 157), the relationship between implicit statistical learning, measured by the Alternating Serial Reaction Time task, and executive functions, measured by multiple neuropsychological tests. In both studies, a modest, but consistent negative correlation between implicit statistical learning and most executive function measures was observed. Factor analysis further revealed that a factor representing the verbal fluency and complex working memory seemed to drive these negative correlations. Thus, an antagonism between implicit statistical learning and executive functions might specifically be mediated by updating component of executive functions or/and long-term memory access.

Keywords: *Executive function, cognitive control, predictive processes, implicit statistical learning, updating, working memory, individual differences*

Introduction

Statistical learning (SL) is a fundamental function of human cognition that allows the implicit extraction of probabilistic regularities from the environment, even without intention, feedback, and reward, and is crucial for predictive processing (Aslin, 2017; Conway, 2020; Kaufman et al., 2010). SL contributes to the acquisition of language (Ullman et al., 2020), motor (Hallgató et al., 2013; Verburgh et al., 2016), musical (Rohrmeier & Rebuschat, 2012; Romano Bergstrom et al., 2012; Shook et al., 2013) and social skills (Baldwin et al., 2008; Parks et al., 2020; Ruffman et al., 2012), as well as habits (Graybiel, 2008; Horváth et al., 2022; Szegedi-Hallgató et al., 2017). SL can occur incidentally, without awareness and the intention to learn (Destrebecqz & Cleeremans, 2001; Fu et al., 2010; Kóbor et al., 2017; Song et al., 2007a, 2007b; Vékony et al., 2022). SL does not function in an isolated manner, but in either cooperative or competitive interactions with other cognitive processes (Poldrack & Packard, 2003, Conway, 2020). Here, we aim to investigate the interaction between implicit SL and executive functions, and to determine which aspects of executive functions show a positive (cooperative) and which a negative (competitive) relationship with statistical learning.

The competition hypothesis was coined in the framework of interactive memory systems (Batterink et al., 2019; Freedberg et al., 2020; Janacsek & Nemeth, 2022; Poldrack & Packard, 2003). According to this framework, learning can rely on either the basal ganglia-based procedural system, or the medial temporal lobe-based declarative system. Larger reliance on one implies a smaller reliance on the other. Initial evidence for this hypothesis has come both from animal (Eichenbaum et al., 1988; Packard et al., 1989) and human neuroimaging studies (Poldrack et al., 1999, 2001). Later, studies rooted in computational neuroscience and reinforcement learning also shed more light on the role of models in guiding learning in these different systems (Daw et al., 2005). It was proposed that the distinction between the declarative and procedural systems might map onto the distinction between model-based and model-free learning algorithms (Doll et al., 2015; Shohamy & Daw, 2014). Model-based learning processes build and make use of a model of the environment to flexibly guide choice, but are more computationally demanding. In contrast, model-free learning processes are computationally less expensive as they rely only on recent outcomes; however, this simplicity allows for less flexibility and sensitivity to contingency changes. The brain is assumed to arbitrate between these two learning systems in a dynamic fashion both during the completion of individual tasks (Doyon et al., 2009; Lee et al., 2014; Poldrack et al., 2001), and also throughout the lifespan (Decker et al., 2016; Nemeth, Janacsek, & Fiser, 2013; Palminteri et al., 2016). Multiple possible arbitration mechanisms have been described, including direct neuroanatomical connections between basal ganglia and medial temporal cortex, differential effects of neuromodulators, and indirect cross-inhibition via prefrontal control processes (Freedberg et al., 2020). The present study focuses on this last mechanism.

The top-down cognitive processes necessary for the flexible cognitive control of behaviour are often collectively referred to as executive functions (EF) (Diamond, 2013; Koechlin, 2016; Miyake & Friedman, 2012). These processes are usually required when we encounter novel, unusual, or constraining situations. They include cognitive functions such as attentional

control, cognitive flexibility, cognitive inhibition and working memory updating (Diamond, 2013; Miyake et al., 2000). Studies exploring the neural basis of EF consistently show that these cognitive processes rely heavily on the prefrontal cortex (PFC) (Badre et al., 2010; Braver et al., 2003; Koechlin et al., 2003; Miller & Cohen, 2001). Moreover, individual differences in EF ability have been associated with individual differences in local prefrontal neural activity (Osaka et al., 2003), as well as functional connectivity between PFC and other brain regions (Kondo et al., 2004). Here, we focus on the question of whether and how individual differences in executive functions might affect statistical learning.

Prefrontal EF has been implicated in arbitrating between learning systems by multiple studies. Lee et al. (2014), for example, showed that lateral PFC and frontopolar cortex seem to encode the reliability associated with both model-based and model-free learning systems, as well as the output of the arbitration process. Importantly, their functional connectivity results also suggested that the arbitration mechanism might work primarily by suppressing model-free learning, when it deems model-based learning to be more beneficial. An intriguing possibility is that the procedural, model-free system is the 'default' learner that is overridden by PFC control involvement. This hypothesis seems to be in line with a series of results that show a negative relationship between statistical learning and prefrontal lobe control processes at both the behavioural and the neural level. For instance, disruption of PFC function by transcranial magnetic stimulation (Ambrus et al., 2020; Smalle et al., 2017), by hypnosis (Nemeth, Janacsek, Polner, et al., 2013), or by cognitive fatigue (Borragán et al., 2016) and engagement of prefrontal lobe resources using dual task conditions (Filoteo et al., 2010; Smalle et al., 2022) all have been reported to increase statistical learning performance. Moreover, multiple neuroimaging studies have also revealed that statistical learning seems to be associated with generally decreased functional connectivity both within PFC circuits and between the PFC and other networks (Park et al., 2022; Tóth et al., 2017). Our goal is to test whether weaker PFC-dependent executive functions could lead to better statistical learning across individuals. That is, do people with relatively weaker executive functions have relatively better statistical learning ability?

The little interindividual differences research carried out so far has suggested that PFC-dependent cognitive functions might relate negatively to statistical learning capability. For instance, higher EF ability has been found to be negatively related to SL ability across individuals (Virag et al., 2015). However, several studies have indicated that working memory is independent from statistical learning (reviewed in Janacsek & Nemeth, 2013). Furthermore, some recent studies have found positive associations between statistical learning and EF ability (Park et al., 2020; Petok et al., 2022). Therefore, the relationship between SL and PFC-supported cognitive functions needs further examination in order to disentangle the still somewhat puzzling relation between the two neurocognitive mechanisms (Janacsek & Németh, 2015). The relationship between implicit statistical learning and PFC-supported cognitive functions has been empirically explored in two studies (Park et al. 2020; Virag et al., 2015); however, both studies had relatively low sample sizes (22 and 40, respectively) leading to low statistical power (Button et al., 2013), and inability to establish reliable correlations between the tasks (Schönbrodt & Perugini, 2013). As neither study included a replication sample, the

robustness of their results is also currently unknown. Furthermore, they only studied the relationship between EF and SL at the task level, without considering the possibility that a pattern might instead emerge at the level of latent EF components, tapped into by multiple tasks. Such latent variable approaches might also lead to relatively higher reliability estimates of cognitive abilities, which is extremely important in inter-individual differences research (Haines et al., 2023). However, the latent structure of EF abilities is itself a contentious issue, with a recent large-scale analysis failing to find measurement models that consistently showed a good fit (Karr et al., 2018).

In this study, we aimed to investigate the relationship between EF and SL, using two large, independent samples acquired in two different studies, offering an internal replication and a far larger overall sample size than previous studies. Both studies measured implicit SL ability using the alternating serial reaction time task (ASRT), a valid and reliable task of implicit SL (Buffington et al., 2021; Farkas et al., 2023), and EF ability using a wide variety of well-characterized neuropsychological tasks. Reasoning in terms of competitive neurocognitive systems, we hypothesized that a negative correlation would be found between EF ability and SL performance. However, due to the heterogeneous findings in the literature regarding SL – EF relationships and the structure of EF itself, we did not formulate strong hypotheses regarding specific tasks and latent EF components. Instead, we focused on assessing EF – SL relationships in a data-driven manner, at both the level of individual EF tasks, as well as the level of latent EF components, tapped into by multiple tasks assessing EF, extracted by exploratory factor analysis.

Methods

Participants

For Study 1, conducted in France, participants took part in a two-day experiment. One hundred eighty nine healthy young adults were included with the following criteria: participants were right-handed, aged under 35 years, with musical training (as measured by practice) inferior to ten years, declared not having active neurological or psychiatric conditions, and declared not be taking any psychoactive medication. Among the 189 participants, two did not come back for the second session and one did not comply with the task instructions on the first session. Thus, the data from the remaining 186 subjects is presented in this study. All participants provided signed informed consent agreements and received financial compensation for their participation. The relevant institutional review board (i.e., the “Comité de Protection des Personnes, CPP Est I” ID: RCB 2019-A02510-57) gave ethical approval for the study.

In Study 2, conducted in Hungary, 180 participants took part in a multi-session experiment. Criteria were the same as in Study 1, with the exception of being under 35 and having no musical expertise. From this pool, we excluded 23 subjects who had missing data on any of the EF tasks. This study was approved by the United Ethical Review Committee for Research in Psychology (EPKEB) in Hungary (Approval number: 30/2012) and by the research ethics

committee of Eötvös Loránd University, Budapest, Hungary. Descriptive statistics of both samples are presented in Table 1.

		Study 1 N = 186	Study 2 N = 157
Age in years (Mean, SD)		22.06 (2.78)	21.80 (4.33)
Gender (N, Percentage)			
	Male	78 (41.9 %)	24 (15.3 %)
	Female	108 (58.1 %)	133 (84.7 %)

Table 1. Descriptive statistics for the two samples.

Tasks

Measure of statistical learning: Alternating serial reaction time (ASRT) task

In Study 1, implicit statistical learning was measured by a modified version of the ASRT task (Howard & Howard, 1997; Nemeth et al., 2010) (Figure 1a). In this task, participants were presented with a yellow arrow on the center of the screen pointing in one of four possible directions (left, up, down or right) for 200 ms. The presentation of the arrow was followed by a presentation of a fixation cross for 500 ms. Using a four-button Cedrus RB-530 response box, participants were instructed to press, as quickly as possible, the button corresponding to the direction of the arrow. Finger placement on each of the buttons of the response box was as follows: the up button had to be pressed with the left index finger; the down button had to be pressed with the right thumb; the right button had to be pressed with the right index; and the left button had to be pressed with left thumb. If participants responded correctly, the fixation cross would remain in the screen for another 750 ms. If participants did not answer or answered incorrectly, an exclamation mark or an “X” would appear for 500 ms, respectively, followed by a 250 ms fixation cross.

Unknowingly to the participants, the appearance of the stimuli followed a predetermined structure where pattern elements (P) alternated with random elements (R) (e.g. 2-R-4-R-3-R-1-R, where the numbers represent the predetermined position of the stimuli, and "R" represent a random position) (Figure 1c &1d). Due to the pattern elements (P) alternating with random elements (R) in the ASRT task, some triplets (runs of three trials) had a higher probability of occurrence than others (Figure 1e). For example, in an 2-R-4-R-3-R-1-R sequence, the (2-3-4) or (3-4-1) triplets have a higher probability of occurrence compared to (2-3-2) or (4-3-1) triplets, as the former type of triplets can be found in P-R-P or in R-P-R structures whereas the second type of triplets can only occur in R-P-R structures. As a result, the former type of triplets are five times more likely to occur than the second type of triplets. Thus, the two types of triplets are referred to as high- and low-probability triplets, respectively. Previous studies using the ASRT task have consistently shown that, with increasing practice in the task, participants' responses to the last element of high-probability triplets become faster compared to responses to the last element of low-probability triplets (Janacsek et al., 2012; Kóbor et al., 2017; Nemeth et al., 2010). Importantly, this phenomenon occurs without the emergence of an explicit knowledge of the sequence structure as reported by the participants (Janacsek et al., 2012; Kóbor et al., 2017; Vékony et al., 2022). Thus, implicit statistical learning in the task can be measured by computing the difference of reaction times between the last elements of high-probability triplets and the last element of low-probability triplets.

This first session consisted of 25 blocks of the ASRT task. There were 85 stimuli in each block, of which the first five were randomly ordered for practice purposes followed by 10 repetitions of the eight-element alternating sequence.

The ASRT task employed in Study 2 had four notable differences to the version used in Study 1. Firstly, in this version of the task, the stimuli (here, a drawing of a dog's head instead of arrows) could appear in one of four horizontally arranged empty circles on the screen, instead of the four cardinal directions (Figure 1b). Secondly, the responses corresponded to the Z, C, B, and M keys on a QWERTY keyboard (with the rest of the keys removed). Thirdly, the task consisted of 45 blocks. We previously observed that although acceptable levels of reliability emerge even with 25 blocks, longer tasks lead to more reliable learning scores, which might be crucial for the correlational analyses we planned here (Farkas et al., 2022). Finally, the timing of the stimuli differed. The task in Study 2 was self-paced (i.e., the target stimulus remained on the screen until the correct response key was pressed) with a response-to-stimulus interval of 120 ms. Despite the differences, as detailed below, robust learning was observed in both studies, with similar distribution of learning scores, except an overall lower mean learning score in Study 1.

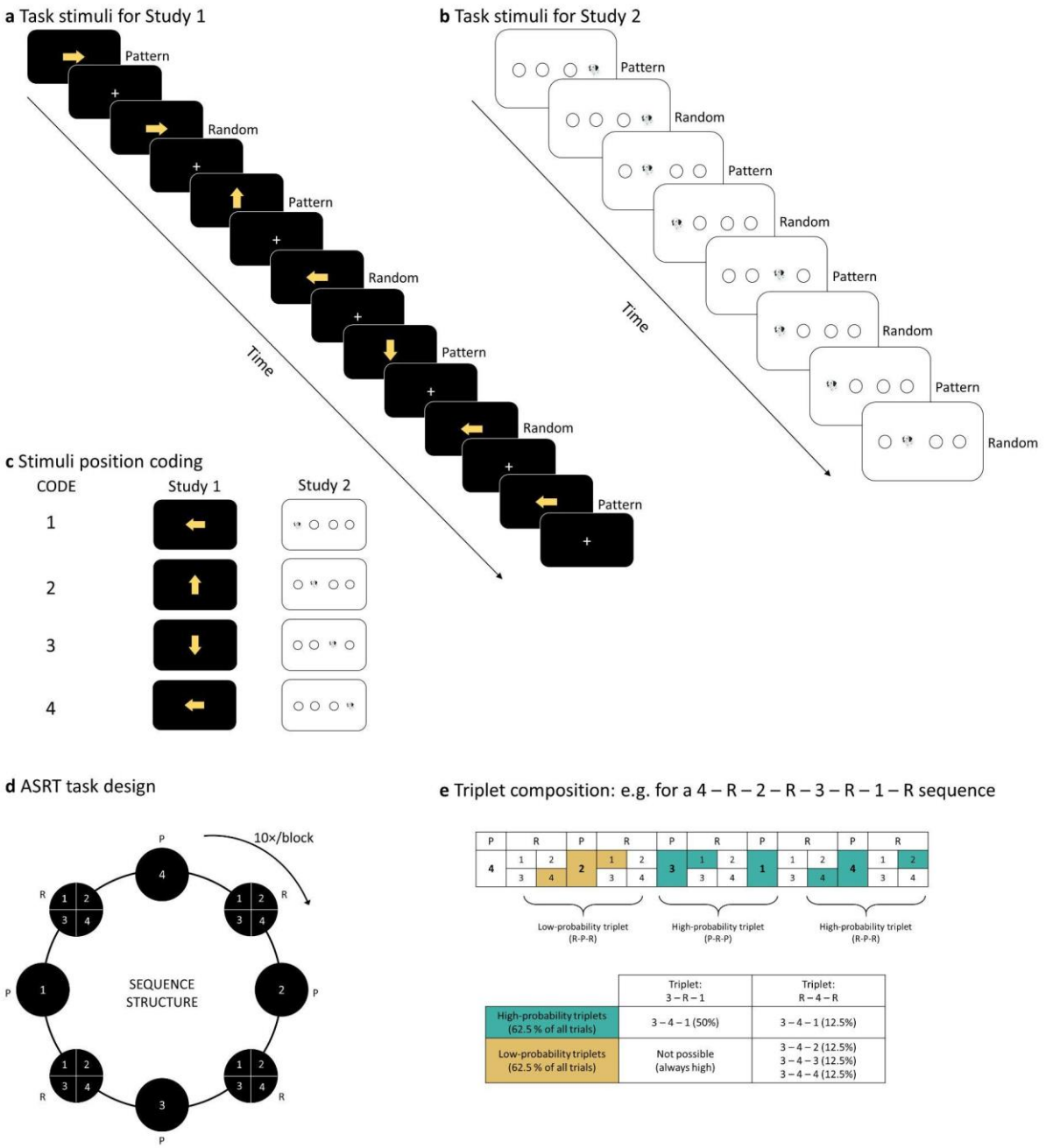

Figure 1. ASRT task design. **a)** In Study 1, the task stimuli consisted of yellow arrows that pointed in one of the four cardinal directions. A fixation cross was presented between each arrow. **b)** In Study 2, task stimuli were the head of a dog that appeared in one out of four different positions. **c)** Each stimuli position can be coded with a number. Here we have 1=left, 2=up, 3=down, 4=right for the arrows in Study 1, and 1=left, 2=center-left, 3=center-right, 4=right of the screen for the dog's head in Study 2. **d)** The stimulus presentation followed an eight-element sequence, in which pattern (P) and random (R) stimuli alternated. The sequence was presented a total of 10 times per block. **e)** Sixty-four different triplets (runs of three consecutive stimuli) could result from the sequence structure. Some of the triplets appear more often than others. High-probability triplets could end in pattern or random element whereas low-probability triplets always ended with a random element. High- and low-probability triplets are denoted in green and yellow, respectively.

Neuropsychological tests for Study 1

Attentional network test (ANT)

The ANT allowed us to measure the capacities of three distinct networks of attention: the alerting, orienting and executive networks (Fan et al., 2002). This task required participants to determine, as fast as possible, whether a central arrow, presented in a set of five arrows, points to the left or to the right. The set of arrows can appear above or below a fixation cross and can be preceded or not by a spatial cue indicating their following location. Furthermore, the central arrow can be congruent (pointing in the same direction) or incongruent (pointing in a different direction) to the other arrows. We followed the standard calculation of network scores (Fan et al., 2002). The alerting component of attention was calculated by subtracting the mean RT of the central cue conditions from the mean of the no-cue conditions for each participant. In a similar manner, the orienting component of attention was calculated by subtracting the mean RT of the spatial cue conditions to the mean RT the center cue conditions and the executive component of attention was calculated by subtracting the mean RT of all congruent conditions to the mean RT of all incongruent conditions. In this task, higher scores for the alerting, orienting and executive scores indicate better attentional performances in these three aspects of attention.

Berg card sorting task (BCST)

We measured set shifting or cognitive flexibility using the computerised version of the BCST.64 available in the Psychology Experiment Building Language (PEBL) software (Mueller & Piper, 2014). In this test, a set of four cards on the top of the screen are presented to the participant. Each card has three characteristics: the colour, the shape and the number of items on the card. The participant was told to match new cards to the cards on the top of the screen according to one of the three characteristics, but they were not told which one; however, they received feedback about whether each choice they made was right or wrong. Thus, the participant was required to find the correct rule (matching characteristic) needed to match the cards as quickly and as accurately as possible. The participant was informed that the rule could change during the task. Cognitive flexibility in the BCST was measured by counting the perseverative errors, meaning the amount of errors reflecting lack of adaptation following a rule switch.

Counting Span (CSPAN) task

Updating or working memory capacity was measured by the CSPAN (Case et al., 1982). In this task, different shapes (blue circles, blue squares, and yellow circles) appeared on the computer screen. The participants' task was to count out loud and retain the amount of blue circles (targets) among the other shapes (distractors) in a series of images. Each image included three to nine blue circles, one to nine blue squares and one to five yellow circles. At the end of each trial, participants repeated orally the total number of targets presented in the image and if counting was correct, the experimenter passed to the following trial When presented with a recall cue at the end of a set, participants had to rename the total number of targets of each

image in their order of presentation. If the recall was correct, participants started a new set containing an extra image. The number of items presented in each image ranged from two to six. When participants made a mistake in the recall, the task was stopped and a new run, starting from a set with two trials, would start again. Each participant completed three runs of the task. Memory span capacity was computed as the mean of the highest set size the participant was able to recall correctly in the three runs.

Go No-go (GNG) task

We measured cognitive inhibition with the computerised version of the GNG task available in the Psychology Experiment Building Language (PEBL) software (Mueller & Piper, 2014). In the GNG task, participants were instructed to respond to certain stimuli ("go" stimuli) by clicking on a button as fast as possible and to refrain from clicking on other stimuli ("no-go" stimuli). In this version, participants were presented with a 2×2 array with four blue stars (one in the centre of each square of the array). Every 1500 ms, a stimulus (the letter P or R) would appear for 500 ms in the place of one of the blue stars. For the first half of the task, the letter P would be the "go" stimulus and the letter R would be the "no-go". This rule would be then inverted in the second half of the task. Participants completed 320 trials. The ratio between "go" and "no-go" trials was 80:20, respectively. Cognitive inhibition capacity in the GNG task was measured by the d':

$$d' = Z(hit\ rate) - Z(false\ alarm\ rate)$$

A higher d' indicated better cognitive inhibition.

Verbal fluency tasks

Verbal fluency was measured with three subtasks testing the lexical, semantic and action components of verbal fluency. In the lexical fluency subtask, participants were required to say as many words as possible starting with the letter P. The letter P is often used in the French version of the phonemic fluency (St-Hilaire et al., 2016). In the semantic fluency subtask, participants were required to name animals, and in the action fluency subtask, isolated verbs describing an action realisable by a person. In each subtask, participants were instructed to say as many words as possible within one minute while avoiding word repetitions, words with the same etymological root and proper nouns. Every time the participant disrespected one of these rules, an error would be counted. Each verbal fluency subtask has a score computed by subtracting the amount of errors from the total amount of words produced within one minute. A higher score in each component of verbal fluency indicated a higher verbal fluency capacity.

Neuropsychological tests for Study 2

Counting Span (CSPAN) task

The procedure for measuring working memory capacity with the CSPAN task was identical in studies 1 and 2.

Digit Span (DSPAN) task

Phonological short-term memory capacity was measured using the DSPAN task (Isaacs & Vargha-Khadem, 1989). In this task, participants listened to and then repeated a list of digits enunciated by an experimenter. The task comprised seven levels of difficulty ranging from three to nine items. Each level comprised four different lists. If participants correctly recalled all items for at least three out of the four lists, they were permitted to move up a level. If participants could not recall at least three lists from the four lists correctly, the task ended. Memory span in the DSPAN task was considered to be the level (three to nine) at which the participant was still able to recall three from the four lists correctly. A higher memory span score is indicative of a better phonological short-term memory capacity.

Corsi blocks tapping task (Corsi) task

Visuo-spatial short-term memory capacity was assessed using the Corsi task (Kessels et al., 2000). In this task, nine cubes were placed in front of the participant in a fixed pseudo-random manner. The blocks were labelled with numbers only visible to the experimenter. The experimenter tapped a number of blocks in a specific sequence after which the participant had to tap the same blocks in the same order. Similarly to the DSPAN, the Corsi task comprised seven levels ranging from three to nine items. Four sequences were presented within a level. Memory span in the Corsi task was considered to be the level (three to nine) at which the participant was still able to recall three from the four lists correctly. A higher memory span score is indicative of a better visuo-spatial short-term memory capacity.

Verbal fluency tasks

Verbal fluency in Study 2 was measured with the Hungarian version of the task (Tánczos et al., 2014). Procedure was similar to the one used in Study 1 with two main differences: the lexical fluency was tested using the letter K and the action fluency was not measured in this study.

Procedure

In Study 1 the experiment was organised over two sessions. During the first session, the ASRT task was administered. Participants were informed that the aim of the study was to study how an extended practice affected performance in a simple reaction time task. Therefore, participants were instructed to respond as fast and as accurately as they could. Participants were not given any information about the underlying structure of the task.

Participants completed the ASRT task in a soundproof experimental booth with a computer screen observable through a window and a four button (up, down, right left) response box placed over a table. Prior to the 25 blocks of the learning phase, the participants completed a three-block training phase where stimuli were completely random. Thus, the first three blocks contained no underlying sequence. This ensured that participants correctly understood the task instructions and familiarised themselves with the response keys. At the end of each block, participants received feedback on the screen reporting their accuracy and reaction time on the

elapsed block. This was followed by a 15 second resting period. Participants were then free to choose when to start the next block.

The ANT, BCST, CSPAN, GNG and verbal fluency tasks were administered in the second experimental session. This session took approximately one hour. In order to avoid a possible fatigue effect in a particular task, the order of presentation of each task was randomised over participants.

The procedure for Study 2 was similar. The ASRT task was administered in the first session and CSPAN, DSPAN, Corsi and verbal fluency tasks were administered in a second session.

Statistical analysis

ASRT learning trajectories

The ASRT blocks were compounded into units of five blocks (epochs) in order to facilitate data processing. ASRT task performance was assessed by calculating the median reaction times (RTs) of correct responses for high- and low-probability triplets separately in each epoch. We then computed learning scores for each epoch by subtracting the median RTs of high-probability triplets from the median RTs of low-probability triplets. A greater difference between high- and low-probability trials indicates greater learning. We excluded from the analysis trills (e.g., 2-1-2) and repetitions (e.g., 2-2-2) as participants show pre-existing tendencies to answer faster in these types of triplets. The first five trials (five warm-up random) of each block and trials with RTs below 100 ms were also removed from the analysis.

To evaluate SL, we conducted a repeated measures analyses of variance (ANOVA) with two factors: epoch (1-5) and triplet (high- vs. low-probability). Greenhouse-Geisser epsilon correction was used when Mauchly's test of sphericity indicated that sphericity cannot be assumed. Original df values and corrected p values (if applicable) are reported together with partial eta-squared (η_p^2) as the measure of effect size. For the correlational analyses, a single index of implicit SL score for each participant was obtained by averaging the learning scores across epochs. Previous results from our group have indicated that while the reliability of learning scores is low for individual epochs, averaging across at least five epochs leads to acceptable levels of reliability. Thus, as adequate reliability is crucial for analyses of inter-individual differences, we decided to conduct our tests of the SL – EF relationship on averaged learning scores, instead of including executive functions score as a covariate in a learning trajectory analysis. We also complemented our classical analysis with Bayesian repeated measures ANOVA, conducted in JASP (JASP Team, 2019). Therefore, we additionally report the inclusion Bayes factors (BF) for each factor, obtained from Bayesian model averaging, and calculated from matched models only (see van den Bergh et al., 2020). The inclusion Bayes factor quantifies the change from prior inclusion odds to posterior inclusion odds and can be interpreted as the evidence in the data for including a predictor or factor. More generally, Bayes factors quantify the relative weight of evidence provided by the data for two theories, the null and the alternative hypotheses, H0 and H1 (Dienes, 2014). According to one commonly used classification (Wagenmakers et al., 2011), BF values between 1 and 3 indicate anecdotal

evidence, values between 3 and 10 indicate substantial evidence, and values above 10 indicate strong evidence for H1. Conversely, values between 1/3 and 1 indicate anecdotal, values between 1/10 and 1/3 indicate substantial evidence, and values below 1/10 indicate strong evidence for H0. Values close to 1 do not support either H0 or H1.

Factor analyses

We further aimed to determine the potential latent structure of our sets of EF measures in a data-driven manner. Therefore, we investigated the factor structure of the EF measures in both of our datasets separately, using maximum likelihood exploratory factor analysis (ML EFA) with varimax rotation, as implemented in the psych package in R (Revelle, 2022), with default settings. To aid interpretation, the scores that reflect error percentages were subtracted from 100, so that higher values represent better performance in all variables. To assess the factorability of the data, we utilised 3 complementary approaches (Dziuban & Shirkey, 1974). Firstly, we inspected the off-diagonal elements of the anti-image covariance matrix. If the dataset is appropriate for factor analysis, these elements should be all above .50. Secondly, we computed the Kaiser-Meyer-Olkin (KMO) test of sampling adequacy. The higher the overall KMO index, the more appropriate a factor analytic model is for the data. The original cut-off recommended by (Kaiser, 1970) is .60, but other authors have also suggested .50 (Dziuban & Shirkey, 1974). Finally, we performed Bartlett's test of sphericity, which tests the hypothesis that the sample correlation matrix came from a multivariate normal population in which the variables of interest are independent (Bartlett, 1950). Rejection of the hypothesis is taken as an indication that the data are appropriate for analysis.

We determined the number of factors to extract using Horn's parallel analysis (Horn, 1965). This approach is based on comparing the eigenvalues of factors of the observed data with those of random data from a matrix of the same size. Factors with higher eigenvalues in the observed, than in the random data are kept. We used a more stringent criteria, and compared the observed eigenvalues to the 95th percentile, instead of the mean of the simulated distributions. Furthermore, our use of ML EFA also allowed us to calculate multiple fit indices of the applied factor analytic models. Following the recommendations of Fabrigar et al. (1999) and Hu & Bentler (1998), we chose to focus on the Root Mean Square Error of Approximation (RMSEA) and the Standardized Root Mean Squared Residual (SRMR). According to a commonly used guideline, RMSEA values less than 0.05 constitute good fit, values in the 0.05 to 0.08 range acceptable fit, values in the 0.08 to 0.10 range marginal fit, and values greater than 0.10 poor fit. Similarly, SRMR values below .08 are generally considered indicators of good model fit (Hu & Bentler, 1999).

After we determined the number of factors to extract, participant level factor scores were calculated for all factors, based on Thomson's (1939) regression method. Correlations between these scores and the ASRT were then assessed. For these correlations, we also calculated Bayes factors, using JASP (JASP Team, 2019), with default priors (stretched beta distribution with a width of 1). We tested the alternative hypothesis that the two variables are negatively correlated. We also tested the robustness of our Bayes factors to different prior widths.

Meta-analysis

In addition to testing our hypotheses in each of the two samples separately, following the recommendation of Braver et al. (2014), we also carried out a continuously cumulating meta-analysis (CCMA) of the two studies. A meta-analytic approach allows us to pool individual effects sizes into a single estimate, while quantifying their heterogeneity. In order to make the two studies comparable, in the two datasets separately, we extracted factor scores from a single factor ML EFA applied to the set of EF tasks that were shared in both studies. These were Lexical fluency, Semantic fluency and CSPAN. We then calculated the Pearson's correlation between these EF factor scores and the implicit SL learning scores and ran our meta-analytic model on these effects. We fit a fixed-effect meta-analysis model implemented in the 'meta' R package (Harrer et al., 2021; Schwarzer, 2022), using the inverse variance pooling method. We expected little between-study heterogeneity in the effect, therefore a priori, we decided on a fixed-effect model to test the average true effect in the two studies. We nevertheless estimated between-study heterogeneity using the REML estimator (Viechtbauer, 2005). As reported below, commonly used measures of heterogeneity indicated little between-studies variability, validating our choice of a fixed-effect model. We relied on the REML estimate of the variance of the distribution of true effects, $\tau 2$, on the percentage of variability in the effect sizes that is not caused by sampling error, I2 (Higgins & Thompson, 2002), and on Cochran's Q (Cochran, 1954), which can be used to test whether there is more variation than can be expected from sampling error alone.

Results

Descriptive statistics of the samples are presented in Table 1. Histograms showing the distributions of all variables are presented in Supplementary Figure S2 and S3, for Study 1 and 2, respectively.

Implicit statistical learning

Study 1

To investigate implicit SL in the ASRT task, we analysed median RTs by a repeated measures ANOVA with two factors: EPOCH (1-5) and TRIPLET (high- vs. low-probability). The main effect of EPOCH was significant ($F_{(4, 740)} = 27.264$, $p < .001$, $\eta_p^2 = .128$, $BF_{inclusion} = 1.154e+18$). indicating general skill learning, i.e., participants became faster overall as the epochs went on, irrespective of triplet probabilities (Figure 2a). The main effect of TRIPLET was significant ($F_{(1, 185)} = 87.855$, $p < .001$, $\eta_p^2 = .322$, $BF_{inclusion} = 7.013e+13$), with average responses to high-probability triplets being faster than responses to low-probability triplets, revealing successful SL. The EPOCH x TRIPLET interaction was also significant ($F_{(4, 740)} = 5.027$, $p < .001$, $\eta_p^2 = .026$, $BF_{inclusion} = 12.291$). Post-hoc tests revealed significantly faster responses to high- as compared to low-probability triplets in all epochs (all $p_{bonf} < .012$), except in epoch 2 ($p_{bonf} = .084$), with progressively larger effect sizes, indicating that SL improved over time as the task went on (also see Figure 2a).

Study 2

We carried out the same repeated measures ANOVA for Study 2 (the sole difference being that the EPOCH factor now had 9 levels), and found similar learning trajectories (Figure 2b). The main effect of EPOCH was significant ($F_{(8, 1248)} = 372.518$, $p < .001$, $\eta_p^2 = .705$, $BF_{inclusion} = \infty$) indicating general skill learning, i.e., participants became faster overall as the epochs went on, irrespective of triplet probabilities. The main effect of TRIPLET was significant ($F_{(1, 156)} = 540.716$, $p < .001$, $\eta_p^2 = .776$, $BF_{inclusion} = 7.386e+48$), with average responses to high-probability triplets being faster than responses to low-probability triplets, revealing successful SL. The EPOCH × TRIPLET interaction was also significant ($F_{(8, 1248)} = 35.405$, $p < .001$, $\eta_p^2 = .185$, $BF_{inclusion} = 1.989e+4$). Post-hoc tests revealed significantly faster responses to high- as compared to low-probability triplets in all epochs (all $p_{bonf} < .026$), with progressively larger effect sizes, indicating that SL improved with practice as the task went on (also see Figure 2b).

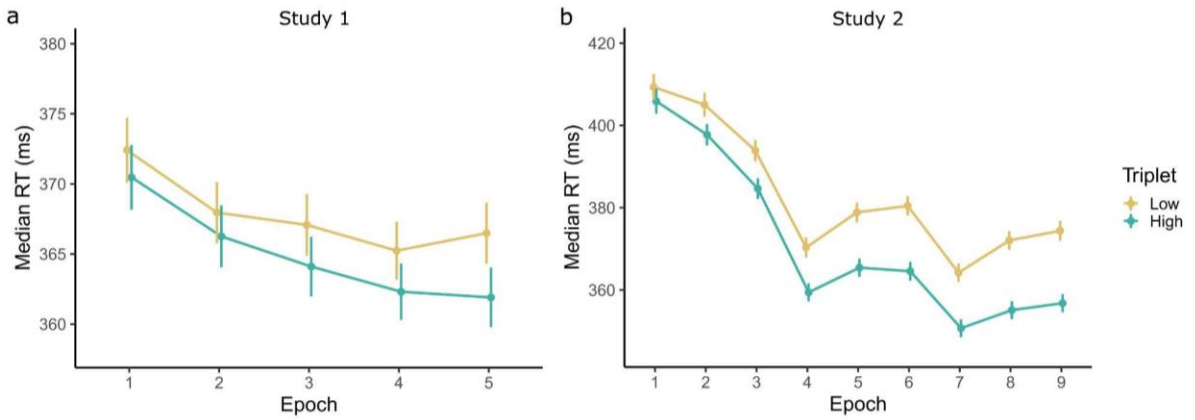

Figure 2. Epoch-wise median reaction times (RTs) in the ASRT tasks for high- and low-probability triplets. **a)** In Study 1 and **b)** in Study 2. There was a significant RT difference between high and low triplets showing statistical learning. SL improved significantly over time. Error bars indicate standard error of the mean. Note the different scales of the y axis.

Correlations between procedural learning and executive functions

Study 1

Bivariate Pearson's correlations between all variables are presented in Figure 3a. EF measures tended to be weakly or moderately positively correlated, the strongest relationships seemed to be between the fluency measures, and between the fluency measures and CSPAN.

Implicit SL ability had generally negative correlations with EF measures, the strongest of these was with Action fluency (Pearson's r = -.165, 95% CI = [-.302, -.021], p = .025, Speaman's rho = -.162, 95% CI = [-.299, -.018], p = .027).

Study 2

Bivariate Pearson's correlations between all variables of Study 2 are presented in Figure 3b. EF measures tended to be weakly or moderately positively correlated, again the fluency measures correlated the strongest.

Implicit SL ability had generally negative correlations with EF measures, the strongest of these was with Semantic fluency (Pearson's r = -.167, 95% CI = [-.315, -.011], *p* = .037, Spearman's rho = -.145, 95% CI = [-.295, .012], *p* = .070).

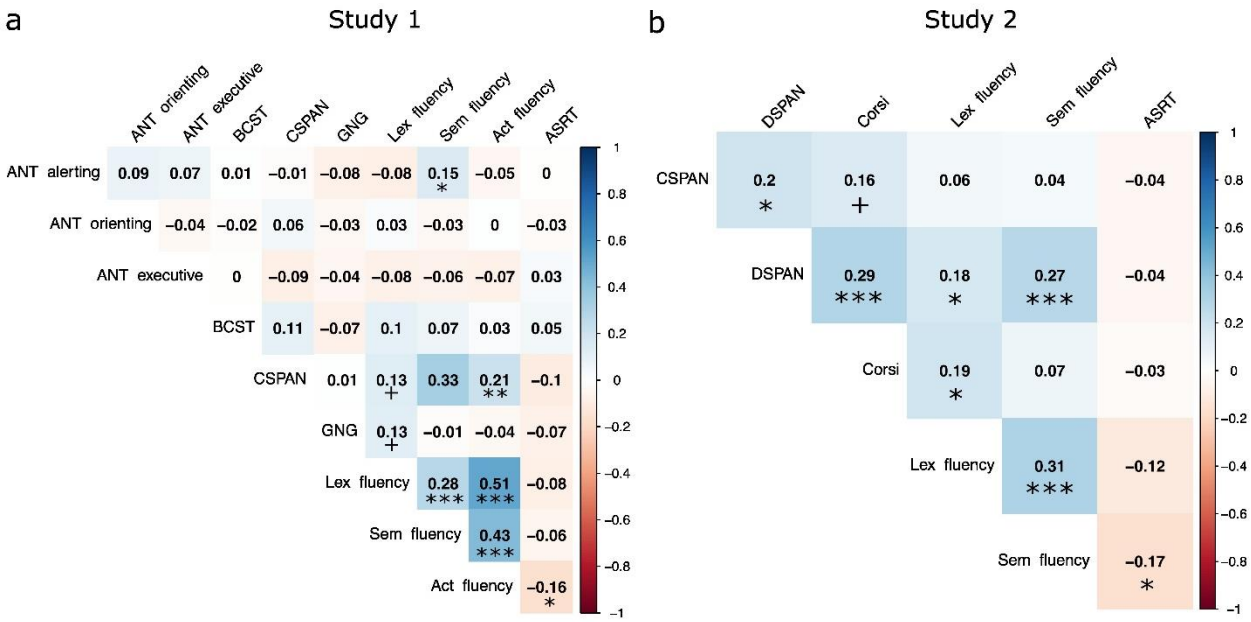

Figure 3. Bivariate correlations between the EF measures and the ASRT. **a)** In Study 1 and **b)** in Study 2. More negative and more positive correlations are indicated by red and blue backgrounds, respectively. The working memory and verbal fluency measures correlated moderately positively with each other, and weakly negatively with ASRT.

***p<.001, **p<.01, *p<.05, +p<.10

Factor analyses of executive function measures

Besides investigating the relationship between performance on individual EF tasks and implicit SL, we also aimed to investigate the relationship between implicit SL and general EF ability that might be captured by shared variance on all tasks. To extract such a common EF measure from the individual tasks, we used Maximum Likelihood Exploratory Factor Analysis (ML EFA). We opted for EFA in order to find a set of latent constructs that capture our specific set of EF measures, while remaining a priori agnostic about the exact structure of these components. While the ML approach allows the computation of various goodness of fit indices, that are unavailable to principal factor approaches.

Study 1

The dataset we considered for factor analysis consisted of the set of EF measures presented in Figure 3a. Beforehand, we established the factorability of the data using 3 approaches. Firstly, the diagonals of the anti-image correlation matrix of the data were all over .5, which suggests good factorability. Secondly, the overall Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy was .59, which is somewhat below the cut-off value of .6 originally suggested by Kaiser (1970), but above the cut-off of .5, suggested by other authors (Dziuban & Shirkey, 1974). Finally, Bartlett's test of sphericity was significant ($\chi2$ (36) = 145.029, p < .001), suggesting adequate factorability. Based on this, the data were deemed appropriate for factor analysis.

To determine the number of factors to extract, we relied on parallel analysis, and goodness of fit indices. Parallel analysis suggested that 1 factor was extractable, based on the comparison of the eigenvalues with randomly generated data (Figure 4a). Thus, the single factor solution was selected.

This factor explained 16% of the variance. The $\chi2$ test of the null hypothesis that one factor is sufficient was not significant ($\chi2$ (27) = 36.87, p = .098), suggesting that the null can be accepted, meaning that our single factor was sufficient in capturing the full dimensionality of the data. Goodness of fit indices were also indicative of good fit (RMSEA = .044, 95% CI = [.000 .077]; SRMR = .062). Factor loadings are presented in Table 2. Factor 1 had high loadings from the three fluency variables and from CSPAN, confirming the pattern of bivariate correlations. We calculated factor scores for each subject, and investigated the correlations with implicit SL.

	Factor 1	Communality
ANT alerting	-.03	.00
ANT orienting	.01	.00
ANT executive	-.11	.01
BCST	.08	.01
CSPAN	**.31**	.10
GNG	.01	.00
Lexical fluency	**.60**	.36
Semantic fluency	**.54**	.29
Action fluency	**.82**	.67

Table 2. Study 1. Factor loadings and communalities of each EF measure on the factors, based on the 1 factor ML solution (16% of variance explained). Loadings above the threshold of .30 are highlighted in bold to aid interpretation.

Factor 1 significantly negatively correlated with implicit SL (Figure 5a, Pearson's r = -.156, 95% CI = [-.293, -.012], *p* = .034, Spearman's rho = - .154, 95% CI = [-.292, -.011], *p* = .036). The Bayes factor for the one-sided alternative hypothesis that the two variables are negatively correlated was BF-0 = 1.683, meaning that the data are approximately 1.7 times more likely to occur under the alternative hypothesis, indicating anecdotal evidence in favour of it. Robustness checks of the Bayes factor to varying prior distribution width are included in the

.jasp files in Supplementary Materials. Histograms showing the distribution of the factor score, as well as the original EF measures are similarly presented in the supplementary material.

Study 2

We proceeded in the same manner for Study 2. The dataset we considered for factor analysis consisted of the set of EF measures presented in Figure 3b. We followed the same approach to establishing the factorability of data, as in the analysis of Study 1. The diagonals of the anti-image correlation matrix of the data were all over .5. The overall KMO measure of sampling adequacy was .60, meeting both more conservative and more liberal cut-off criteria. Finally, Bartlett's test of sphericity was significant ($\chi2$ (10) = 54.591, $p < .001$). Based on this, the data were deemed appropriate for factor analysis.

To determine the number of factors to extract, we again relied on parallel analysis. Parallel analysis suggested that 2 factors are extractable, based on the comparison of their eigenvalues with random data (Figure 4b). Thus, the 2 factor solution was selected.

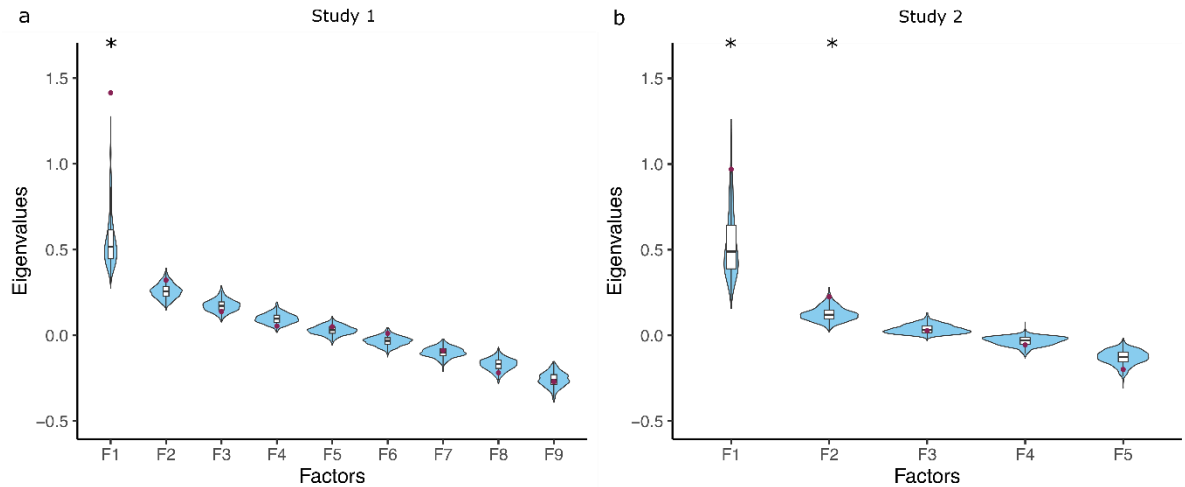

Figure 4. Parallel analysis to determine the extractable factors. **a)** In Study 1 and **b**) in Study 2. The observed eigenvalues are indicated by red dots. These were compared to the distribution of eigenvalues obtained from simulated data, indicated by violin plots and boxplots. Stars indicate factors for which the observed eigenvalue is larger than the 95th percentile of the simulated distributions, and thus were deemed extractable.

These two factors explained 19% and 14% of the variance, respectively. The total variance explained of the model was thus 34%, somewhat more than the explained variance of the Study 1 EFA model. The $\chi2$ test of the null hypothesis that 2 factors are sufficient was not significant ($\chi2$ (1) = 1.40, p = .236), suggesting that the null can be accepted, meaning that our 2 factors are sufficient in capturing the full dimensionality of the data. Goodness of fit indices were also indicative of good fit (RMSEA = .050, 95% CI = [.000 .227]; SRMR = .020), although the RMSEA had a noticeably larger CI, compared to the Study 1 EFA model. Factor loadings are presented in Table 3. Factor 1 had high loadings from the 2 fluency variables and a somewhat weaker loading from DSPAN. Factor 2 had high positive loadings from the CSPAN, DSPAN

and Corsi tasks. Thus, the factor structure reflected the dissociation between the fluency and the short-term memory measures.

	Factor 1	Factor 2	Communality
CSPAN	.03	**.32**	.10
DSPAN	.28	**.52**	.34
Corsi	.08	**.54**	.30
Lexical fluency	**.33**	.23	.16
Semantic fluency	**.88**	.03	.77

Table 3. Study 2. Factor loadings and communalities of each EF measure on the 2 factors, based on the 2 factor varimax rotated ML solution (34% of variance explained). Loadings above the threshold of .30 are highlighted in bold to aid interpretation.

Factor 1 significantly negatively correlated with implicit SL, and with a similar effect size, as in Study 1 (Figure 5b, Pearson's r = -.170, 95% CI = [-.319, -.014], *p* = .033, Spearman's rho = -.150, 95% CI = [-.300, .007], *p* = .060). The Bayes factor for the one-sided alternative hypothesis that the two variables are negatively correlated was BF-0 = 1.875, meaning that the data are approximately 1.9 times more likely to occur under the alternative hypothesis, indicating anecdotal evidence in favour of it. Factor 2 did not correlate significantly with implicit SL ability (Factor 2: Pearson's r = -.034, 95% CI = [-.190, .123], *p* = .672, Spearman's rho = -.023, 95% CI = [-.179, .134], *p* = .771, BF-0 = 0.145). Robustness checks of the Bayes factor to varying prior distribution width are included in the .jasp files in Supplementary Materials. Histograms showing the distribution of the 2 factor scores, as well as the original EF measures are similarly presented in the supplementary material.

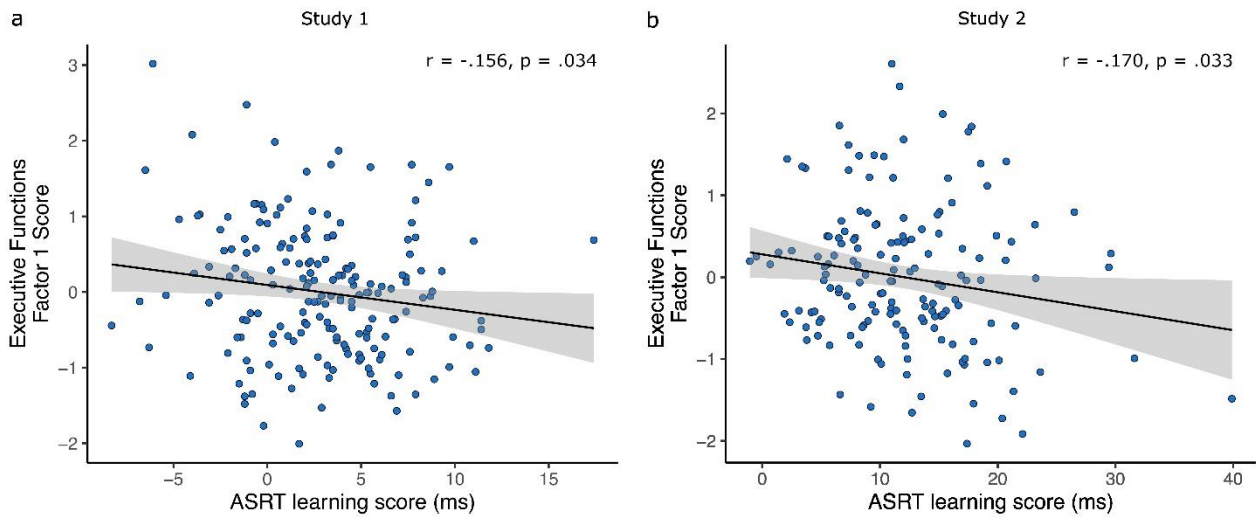

Figure 5. Relationship between ASRT learning scores and EF Factor 1 scores across subjects. Solid line is the linear fit, shaded area corresponds to the 95% CI. One datapoint per participant. **a)** In Study 1 and **b**) in Study 2.

Continuously Cumulating Meta-Analysis

Besides testing the hypothesis of a negative association between common EF ability and implicit SL in our two samples separately, we also ran a fixed-effect meta-analytic model, in order to pool evidence from both of our studies into a single effect size, while estimating their heterogeneity. Heterogeneity metrics revealed little between-studies variability in the effect, validating our choice of a fixed-effect model. The between-study heterogeneity variance was estimated at $\tau^2 = 0$, with an I^2 value of 0%, Cochran's Q test was also not statistically significant, Q(1) = 0.60, p = .440. The pooled effect size was negative and significantly different from zero, r = -.115, 95% CI = [-.218, -.008], *p* = .035 (Figure 6). The weights of Study 1 and Study 2 in the pooled effect were 54.3% and 45.7%, respectively. We also carried out this analysis using factor scores derived from all EF measures in both tasks, instead of just the shared ones (Supplementary Figure S1). This analysis also yielded a pooled effect size, significantly more negative than 0, r = -.162, 95% CI = [-.264, -.057], *p* = .003.

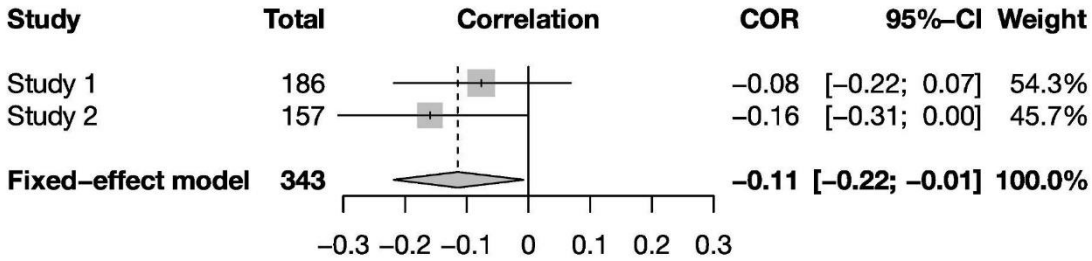

Figure 6. CCMA of the two studies. Pearson's r between EF factor scores and implicit SL learning scores, along with its 95% CI is shown next to each individual study. Below, the pooled effect size from a fixed-effect meta-analytic model and its 95% CI is shown. Individual and total sample sizes are also indicated, as well as study weight. Factor scores were derived only from tasks that were shared by both studies.

Discussion

In this study, we explored how implicit SL relates to multiple components of prefrontal lobe-dependent EF. At the level of individual tasks, SL correlated negatively with most EF measures, the strongest of these correlations being with verbal fluency. At the level of latent EF components, an EF factor comprising verbal fluency and counting span scores in Study 1 and verbal fluency and digit span scores in Study 2 also correlated negatively with SL, with a modest, but similar effect size in both studies. Our results imply that individuals with better verbal fluency and working memory ability have lower implicit statistical learning ability, suggesting that specific prefrontal lobe functions might interfere with implicit statistical learning while still allowing for the acquisition of the underlying regularities. We aimed to improve upon previous studies and extend their results in several ways. We made use of far larger samples ($N_{Study1} = 186$, $N_{Study2} = 157$), giving us larger statistical power than any previous study in the field (Button et al., 2013). Our successful internal replication, using a second sample and a cumulative meta-analytic approach, speaks to the robustness of our results (Braver et al., 2014). Our measurement of EF components at both the task and the latent variable level using exploratory factor analysis allowed for a more thorough characterization of this complex cognitive function (Diamond, 2013), and attenuated measurement error (Haines et al., 2023), while remaining agnostic about the measurement model of EF (Karr et al., 2018). Finally, we used an implicit SL task that has been shown to be both valid (Buffington et al., 2021) and reliable (Farkas et al., 2023), which is crucial, given the recent debate about the psychometric properties of commonly used implicit learning tasks (West et al., 2018).

We interpret our results in the PFC-mediated competition model of declarative-procedural interactions (Freedberg et al., 2020; Janacsek & Nemeth, 2022; Poldrack & Packard, 2003), and suggest that the observed negative relationship between EF and SL might be due to the suppression of model-free, procedural learning by prefrontal EF. Support for this theory has come not only from the effects of PFC disruption on learning (Ambrus et al., 2020; Borragán et al., 2016; Filoteo et al., 2010; Nemeth et al., 2013; Smalle et al., 2017, 2022) and from computational modelling (Lee et al., 2014) studies, but also from neural data. For example, a meta-analysis of multiple primate studies by Loonis et al. (2017) has revealed distinct patterns of post-choice oscillatory synchrony within PFC during implicit versus explicit learning, such that Delta/Theta band synchrony increased after correct choices during implicit learning, but after incorrect choices during explicit learning. Moreover, their results also suggested that whereas explicit learning was associated with increased synchrony between PFC and hippocampus in the alpha and beta bands, implicit learning was associated with decreased synchrony between PFC and caudate in the theta band. A similar pattern of results was revealed in humans by Voss et al. (2012), who showed that the use of a flexible, declarative learning strategy was linked to the interaction between the medial temporal lobe and the fronto-parietal attentional network, whereas the use of a more rigid, procedural learning strategy was linked to caudate nucleus fronto-parietal network interactions. Furthermore, neuroimaging studies also indicate that statistical learning is associated with decreased functional connectivity within PFC circuits and between the PFC and other networks (Park et al., 2022; Tóth et al., 2017).

In Study 1, the most parsimonious model consisted of a single factor, comprising fluency and counting span performances. The link between verbal fluency and working memory tasks is not surprising. Fluency scores have constantly been found to relate to verbal working memory (Daneman, 1991; Rende et al., 2002; Unsworth et al., 2011; Unsworth & Engle, 2005; Woods et al., 2005). Higher working memory capacity may aid in handling the cognitive load in fluency tasks (Rosen & Engle, 1997), while lower capacity may lead to more perseveration errors (Azuma, 2004). Shao et al. (2014) have also found that both category and letter fluency scores were uniquely predicted by updating in working memory. Similar findings that relate updating to verbal fluency performance were found by Gustavon et al. (2019). The factor structure uncovered by Fisk & Sharp (2004) is also suggestive of this relationship. Although they loaded onto a separate factor more strongly, a factor representing updating also had quite high loadings from word fluency tasks. Thus, the factor in Study 1 may reflect updating. This makes sense when considering the demands of verbal fluency and counting span tasks. Both tasks require tracking and updating sequences of items. In verbal fluency tasks, individuals generate words fitting a specific category, requiring constant updating of working memory. Without updating, they may repeat words or struggle to generate new ones, affecting performance. Counting span tasks involve maintaining and updating a number sequence while performing a secondary task. Without updating, individuals may lose track of the sequence, negatively impacting performance. Updating appears to be a suitable explanation for the common factor in verbal fluency and counting span tasks. Although inhibition and set shifting have been proposed as significant factors in fluency tasks (Shao et al., 2014), the absence of positive correlations between our measures of shifting (BCST) and inhibition (GNG) and other executive function tasks makes it challenging to consider them as appropriate explanations for factor 1.

While our general findings, suggestive of a negative relationship across individuals in EF and SL are in line with the competitive neurocognitive systems framework (Janacsek & Nemeth, 2022; Poldrack & Packard, 2003; Reber, 2013), the fact that it is primarily for the updating component that this association was uncovered runs contrary to some previous results showing a positive relationship between working memory and statistical learning (Frensch & Miner, 1994; Howard & Howard, 1997; Park et al., 2020). The central idea of these studies was that larger working memory capacity might open up a larger “window” for serial order learning. However, as noted by Janacsek and Nemeth (2013), these effects seem primarily observed in explicit learning conditions and consolidation, rather than implicit learning of probabilistic representations per se. Our results imply that under implicit learning, EF updating might instead compete with statistical learning. A possible explanation might be that updating might disrupt the stabilization of a simple predictive model allowing triplet learning in the ASRT task. However, we note that the nature of this kind of “updating” of implicit probabilistic representations is likely quite different from the explicit updating of items in working memory that is involved in EF tasks. Cognitive control, executive functions and working memory are related to a model-based computational strategy (Otto, Gershman, et al., 2013; Otto, Raio, et al., 2013). So our results might be interpretable as showing that more automatic, model-free learning is antagonistically related to model-based updating associated with goal-directed processes.

An alternative explanation to updating is that both working memory and fluency are strongly related to long-term memory access. Numerous theories and empirical evidence support the former (Cowan, 2016; Unsworth et al., 2013), while the fluency test merely involves retrieving words from the mental lexicon (Ullman, 2007). From this, we can speculate that if access to our long-term representations and models is poorer, it significantly enhances model-free learning because there is no interference from previous models in learning new patterns and predictions. This explanation aligns with the explanation provided by Ambrus et al. (2020), who attributed their results of inhibitory TMS-enhanced statistical learning to difficulties in long-term memory access or the suppression of top-down processes. However, this hypothesis needs to be validated through studies that examine long-term memory access, cognitive control, and predictive processes not only through behavioral but also neuroimaging techniques within a single experimental design.

While our study has notable strengths, including the incorporation of two large, independent samples, multiple EF measures, and the use of data-driven latent component extraction, there are some limitations that need to be mentioned, and addressed in future research. Firstly, the effects sizes were relatively small. This likely stems partially from the wide array of factors that influence EF and SL ability, and partially from attenuated correlations due to measurement error. We have recently estimated the reliability of the ASRT task in similar settings as the current study, to be between .754 and .791. If we correct our correlation coefficients for attenuation with the lowest reliability value from that study, we obtain $r = \frac{-.155}{\sqrt{.754}} = -.179$ and $r = \frac{-.170}{\sqrt{.754}} = -.196$, which are likely better estimates of the true relationships. Secondly, to replicate the results obtained in Study 1, we re-analysed data from a sample previously collected by us, and first described in Kóbor et al. (2017). While we note that as this is an earlier study, which was not designed with the research question of this study in mind, we believe the differences between protocols do not meaningfully diminish the informativeness of our replication attempt. Indeed, the two primary differences are the different set of EF tasks, and the longer ASRT. Regarding the first, the results of Study 1 show that tasks tapping into the updating component of EF (CSPAN and verbal fluency) showed the clearest relationship with implicit SL. These are exactly the tasks that were shared between studies. Regarding the second study, we have previously shown that longer ASRT tasks lead to more reliable learning scores (Farkas et al., 2022), which are crucial for correlational research designs. Thus, if anything the longer task should increase the power of our replication attempt. However, the different response to stimulus intervals of the two studies does decrease their comparability. As a result, it would be beneficial for future studies and replication attempts to show a greater consistency across studies in the protocol. Finally and relatedly, the incomplete nature of our EF battery should also be addressed by future work. Our set of EF tasks did not cover the most widely accepted theoretical model of EF components in an equal manner, as for set shifting and inhibition, we only had the BCST and GNG tasks, respectively (Miyake et al., 2000). Thus, our results should be extended and confirmed by future studies with a more complete set of EF tasks, and a more theory-driven assessment of latent EF components, for example with Confirmatory Factor Analytic models. We note however, that the ubiquity of the Unity and Diversity model has been questioned recently (Karr et al., 2018).

To our knowledge this was the first study to combine multiple data sets to investigate the relationship between implicit SL and EF testing large pools of participants. We found that general executive function has a negative relationship with implicit SL, suggesting that individuals with better EF functions might be worse at acquiring probabilistic models of visuospatial information. Our results fall in line with the theoretical framework stating that implicit automatic processing works in competition with PFC functions. This study takes a step further by exploring which components of executive functioning might be at the core of this competition. Our results are indicative that it is the updating component that is driving this relationship. Our study highlights the importance of exploring the relationship between different cognitive abilities using larger sets of subjects and in a data-driven manner.

Credit author statement

Felipe Pedraza: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Data curation, Writing – Original draft, Writing – Review & Editing, Visualization
Bence C. Farkas: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – Original draft, Writing – Review & Editing, Visualization
Teodóra Vékony: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – Review & Editing, Visualization, Supervision, Project administration
Frederic Haesebaert: Inclusion of participants, Writing – Review & Editing
Karolina Janacsek: Conceptualization, Methodology, Investigation, Data curation, Project administration, Funding acquisition
Royce Anders: Writing – Review & Editing
Barbara Tillmann: Writing – Review & Editing
Gaën Plancher: Writing – Review & Editing
Dezso Nemeth: Conceptualization, Methodology, Investigation, Data curation, Writing – Review & Editing, Supervision, Project administration, Funding acquisition

Declarations of competing interest

None.

Open practices

All code and data necessary to replicate the findings of this manuscript can be found on the OSF platform, at the following link:
https://osf.io/2asnb/

Acknowledgements

The authors are grateful to Jeoffrey Maillard for managing participant enrollment. This work was supported by the Chaire de Professeur Junior Program by INSERM and French National Grant Agency (Chaire Inserm-ANR-22-CE64-001); the National Brain Research Program (NAP2022-I-2/2022), János Bolyai Research Scholarship of the Hungarian Academy of Sciences (to K.J.). Project no. 124148 has been implemented with the support provided by the Ministry of Innovation and Technology of Hungary from the National Research, Development, and Innovation Fund, financed under the NKFI/OTKA PD funding scheme (to K.J.).

Bibliography

Ambrus, G. G., Vékony, T., Janacsek, K., Trimborn, A. B. C., Kovács, G., & Nemeth, D. (2020). When less is more: Enhanced statistical learning of non-adjacent dependencies after disruption of bilateral DLPFC. *Journal of Memory and Language*, *114*, 104144. https://doi.org/10.1016/j.jml.2020.104144

Azuma, T. (2004). Working Memory and Perseveration in Verbal Fluency. *Neuropsychology*, *18*, 69–77. https://doi.org/10.1037/0894-4105.18.1.69

Badre, D., Kayser, A. S., & D'Esposito, M. (2010). Frontal Cortex and the Discovery of Abstract Action Rules. *Neuron*, *66*(2), 315–326. https://doi.org/10.1016/j.neuron.2010.03.025

Baldwin, D., Andersson, A., Saffran, J., & Meyer, M. (2008). Segmenting dynamic human action via statistical structure. *Cognition*, *106*(3), 1382–1407. https://doi.org/10.1016/j.cognition.2007.07.005

Bartlett, M. S. (1950). Tests of significance in factor analysis. *British Journal of Psychology*, *3*, 77–85.

Batterink, L. J., Paller, K. A., & Reber, P. J. (2019). Understanding the neural bases of implicit and statistical learning. Topics in cognitive science, 11(3), 482-503.

Borragán, G., Slama, H., Destrebecqz, A., & Peigneux, P. (2016). Cognitive Fatigue Facilitates Procedural Sequence Learning. *Frontiers in Human Neuroscience*, *10*. https://doi.org/10.3389/fnhum.2016.00086

Braver, S. L., Thoemmes, F. J., & Rosenthal, R. (2014). Continuously Cumulating Meta-Analysis and Replicability. *Perspectives on Psychological Science*, *9*(3), 333–342. https://doi.org/10.1177/1745691614529796

Braver, T. S., Reynolds, J. R., & Donaldson, D. I. (2003). Neural Mechanisms of Transient and Sustained Cognitive Control during Task Switching. *Neuron*, *39*(4), 713–726. https://doi.org/10.1016/S0896-6273(03)00466-5

Buffington, J., Demos, A. P., & Morgan-Short, K. (2021). The reliability and validity of procedural memory assessments used in second language acquisition research. *Studies in Second Language Acquisition*, *43*(3), 635–662. https://doi.org/10.1017/S0272263121000127

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*(5), 365–376. https://doi.org/10.1038/nrn3475

Case, R., Kurland, D. M., & Goldberg, J. (1982). Operational efficiency and the growth of short-term memory span. *Journal of Experimental Child Psychology*, *33*(3), 386–404. https://doi.org/10.1016/0022-0965(82)90054-6

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181–204. https://doi.org/10.1017/S0140525X12000477

Cochran, W. G. (1954). Some Methods for Strengthening the Common χ2 Tests. *Biometrics*, *10*(4), 417–451. https://doi.org/10.2307/3001616

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

Conway, C. M., & Christiansen, M. H. (2005). Modality-Constrained Statistical Learning of Tactile, Visual, and Auditory Sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 24–39. https://doi.org/10.1037/0278-7393.31.1.24

Conway, C. M., & Christiansen, M. H. (2016). Statistical Learning Within and Between Modalities: Pitting Abstract Against Stimulus-Specific Representations. *Psychological Science*. https://journals.sagepub.com/doi/10.1111/j.1467-9280.2006.01801.x

Cowan, N. (2016). *Working memory capacity: Classic edition*. Psychology press.

Daneman, M. (1991). Working memory as a predictor of verbal fluency. *Journal of Psycholinguistic Research*, *20*(6), 445–464. https://doi.org/10.1007/BF01067637

Daw, N. D., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience*, *8*(12), 1704–1711. https://doi.org/10.1038/nn1560

Decker, J. H., Otto, A. R., Daw, N. D., & Hartley, C. A. (2016). From Creatures of Habit to Goal-Directed Learners: Tracking the Developmental Emergence of Model-Based Reinforcement Learning. *Psychological Science*, *27*(6), 848–858. https://doi.org/10.1177/0956797616639301

Destrebecqz, A., & Cleeremans, A. (2001). Can sequence learning be implicit? New evidence with the process dissociation procedure. *Psychonomic Bulletin & Review*, *8*(2), 343–350. https://doi.org/10.3758/BF03196171

Diamond, A. (2013). Executive Functions. *Annual Review of Psychology*, *64*(1), 135–168. https://doi.org/10.1146/annurev-psych-113011-143750

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00781

Doll, B. B., Shohamy, D., & Daw, N. D. (2015). Multiple memory systems as substrates for multiple decision systems. *Neurobiology of Learning and Memory*, *117*, 4–13. https://doi.org/10.1016/j.nlm.2014.04.014

Doyon, J., Bellec, P., Amsel, R., Penhune, V., Monchi, O., Carrier, J., Lehéricy, S., & Benali, H. (2009). Contributions of the basal ganglia and functionally related brain structures to motor learning. *Behavioural Brain Research*, *199*(1), 61–75. https://doi.org/10.1016/j.bbr.2008.11.012

Dziuban, C. D., & Shirkey, E. C. (1974). When is a correlation matrix appropriate for factor analysis? Some decision rules. *Psychological Bulletin*, *81*, 358–361. https://doi.org/10.1037/h0036316

Eichenbaum, H., Fagan, A., Mathews, P., & Cohen, N. J. (1988). Hippocampal system dysfunction and odor discrimination learning in rats: Impairment of facilitation depending on representational demands. *Behavioral Neuroscience*, *102*, 331–339. https://doi.org/10.1037/0735-7044.102.3.331

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, *4*, 272–299. https://doi.org/10.1037/1082-989X.4.3.272

Fan, J., McCandliss, B. D., Sommer, T., Raz, A., & Posner, M. I. (2002). Testing the Efficiency and Independence of Attentional Networks. *Journal of Cognitive Neuroscience*, *14*(3), 340–347. https://doi.org/10.1162/089892902317361886

Fanuel, L., Pleche, C., Vékony, T., Janacsek, K., Nemeth, D., & Quentin, R. (2022). How does the length of short rest periods affect implicit probabilistic learning? *Neuroimage: Reports*, 2(1), 100078.

Farkas, B. C., Krajcsi, A., Janacsek, K., & Nemeth, D. (2023). The complexity of measuring reliability in learning tasks: An illustration using the Alternating Serial Reaction Time Task. *Behavior Research Methods*. https://doi.org/10.3758/s13428-022-02038-5

Filoteo, J. V., Lauritzen, S., & Maddox, W. T. (2010). Removing the frontal lobes: The effects of engaging executive functions on perceptual category learning. *Psychological Science*, *21*(3), 415–423. https://doi.org/10.1177/0956797610362646

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

Fisk, J. E., & Sharp, C. A. (2004). Age-Related Impairment in Executive Functioning: Updating, Inhibition, Shifting, and Access. *Journal of Clinical and Experimental Neuropsychology*, *26*(7), 874–890. https://doi.org/10.1080/13803390490510680

Freedberg, M., Toader, A. C., Wassermann, E. M., & Voss, J. L. (2020). Competitive and cooperative interactions between medial temporal and striatal learning systems. *Neuropsychologia*, *136*, 107257. https://doi.org/10.1016/j.neuropsychologia.2019.107257

Frensch, P. A., & Miner, C. S. (1994). Effects of presentation rate and individual differences in short-term memory capacity on an indirect measure of serial learning. *Memory & Cognition*, *22*(1), 95–110. https://doi.org/10.3758/bf03202765

Frost, R., Armstrong, B. C., Siegelman, N., & Christiansen, M. H. (2015). Domain generality versus modality specificity: The paradox of statistical learning. *Trends in Cognitive Sciences*, *19*(3), 117–125. https://doi.org/10.1016/j.tics.2014.12.010

Fu, Q., Dienes, Z., & Fu, X. (2010). Can unconscious knowledge allow control in sequence learning? *Consciousness and Cognition*, *19*(1), 462–474. https://doi.org/10.1016/j.concog.2009.10.001

Graybiel, A. M. (2008). Habits, rituals, and the evaluative brain. *Annual Review of Neuroscience*, *31*, 359–387. https://doi.org/10.1146/annurev.neuro.29.051605.112851

Gustavson, D. E., Panizzon, M. S., Franz, C. E., Reynolds, C. A., Corley, R. P., Hewitt, J. K., Lyons, M. J., Kremen, W. S., & Friedman, N. P. (2019). Integrating verbal fluency with executive functions: Evidence from twin studies in adolescence and middle age. *Journal of Experimental Psychology: General*, *148*, 2104–2119. https://doi.org/10.1037/xge0000589

Haines, N., Sullivan-Toole, H., & Olino, T. (2023). From Classical Methods to Generative Models: Tackling the Unreliability of Neuroscientific Measures in Mental Health Research. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, S2451902223000022. https://doi.org/10.1016/j.bpsc.2023.01.001

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. https://doi.org/10.1016/j.cortex.2012.01.002

Harrer, M., Cuijpers, P., Furukawa, T., & Ebert, D. (2021). *Doing Meta-Analysis with R: A Hands-On Guide*. Chapman and Hall/CRC. https://doi.org/10.1201/9781003107347

Hauser, M. D., Newport, E. L., & Aslin, R. N. (2001). Segmentation of the speech stream in a non-human primate: Statistical learning in cotton-top tamarins. *Cognition*, *78*(3), B53–B64. https://doi.org/10.1016/S0010-0277(00)00132-3

Henke, K. (2010). A model for memory systems based on processing modes rather than consciousness. *Nature Reviews. Neuroscience*, *11*(7), 523–532. https://doi.org/10.1038/nrn2850

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, *21*(11), 1539–1558. https://doi.org/10.1002/sim.1186

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*(2), 179–185. https://doi.org/10.1007/BF02289447

Horváth, K., Nemeth, D., & Janacsek, K. (2022). Inhibitory control hinders habit change. *Scientific Reports*, *12*(1), Article 1. https://doi.org/10.1038/s41598-022-11971-6

Howard, J. H., & Howard, D. V. (1997). Age differences in implicit learning of higher order dependencies in serial patterns. *Psychology and Aging*, *12*(4), 634–656.

Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, *3*, 424–453. https://doi.org/10.1037/1082-989X.3.4.424

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Isaacs, E. B., & Vargha-Khadem, F. (1989). Differential course of development of spatial and verbal memory span: A normative study. *British Journal of Developmental Psychology*, *7*(4), 377–380. https://doi.org/10.1111/j.2044-835X.1989.tb00814.x

Janacsek, K., Fiser, J., & Nemeth, D. (2012). The best time to acquire new skills: Age-related differences in implicit sequence learning across the human lifespan. *Developmental Science*, *15*(4), 496–505. https://doi.org/10.1111/j.1467-7687.2012.01150.x

Janacsek, K., & Nemeth, D. (2012). Predicting the future: From implicit learning to consolidation. *International Journal of Psychophysiology*, *83*(2), 213–221. https://doi.org/10.1016/j.ijpsycho.2011.11.012

Janacsek, K., & Nemeth, D. (2013). Implicit sequence learning and working memory: Correlated or complicated? *Cortex*, *49*(8), 2001–2006. https://doi.org/10.1016/j.cortex.2013.02.012

Janacsek, K., & Németh, D. (2015). The puzzle is complicated: When should working memory be related to implicit sequence learning, and when should it not? (Response to Martini et al.). *Cortex*, *64*, 411–412.

Johansson, T. (2009). Strengthening the Case for Stimulus-Specificity in Artificial Grammar Learning. *Experimental Psychology*, *56*(3), 188–197. https://doi.org/10.1027/1618-3169.56.3.188

Kaiser, H. F. (1970). A second generation little jiffy. *Psychometrika*, *35*, 401–415. https://doi.org/10.1007/BF02291817

Karr, J. E., Areshenkoff, C. N., Rast, P., Hofer, S. M., Iverson, G. L., & Garcia-Barrera, M. A. (2018). The unity and diversity of executive functions: A systematic review and re-

analysis of latent variable studies. *Psychological Bulletin*, *144*(11), 1147–1185. https://doi.org/10.1037/bul0000160

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

Kessels, R. P. C., van Zandvoort, M. J. E., Postma, A., Kappelle, L. J., & de Haan, E. H. F. (2000). The Corsi Block-Tapping Task: Standardization and Normative Data. *Applied Neuropsychology*, *7*(4), 252–258. https://doi.org/10.1207/S15324826AN0704_8

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition*, *83*(2), B35–B42. https://doi.org/10.1016/S0010-0277(02)00004-5

Kóbor, A., Janacsek, K., Hermann, P., Zavecz, Z., Varga, V., Csépe, V., … Nemeth, D. (2022, January 12). Predictability-dependent encoding of statistical regularities in the early visual cortex. https://doi.org/10.31234/osf.io/axq49

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

Koechlin, E. (2016). Prefrontal executive function and adaptive behavior in complex environments. *Current Opinion in Neurobiology*, *37*, 1–6. https://doi.org/10.1016/j.conb.2015.11.004

Koechlin, E., Ody, C., & Kouneiher, F. (2003). The Architecture of Cognitive Control in the Human Prefrontal Cortex. *Science*, *302*(5648), 1181–1185. https://doi.org/10.1126/science.1088545

Kolodny, O., Edelman, S., & Lotem, A. (2014). The evolution of continuous learning of the structure of the environment. *Journal of The Royal Society Interface*, *11*(92), 20131091. https://doi.org/10.1098/rsif.2013.1091

Kondo, H., Osaka, N., & Osaka, M. (2004). Cooperation of the anterior cingulate cortex and dorsolateral prefrontal cortex for attention shifting. *NeuroImage*, *23*(2), 670–679. https://doi.org/10.1016/j.neuroimage.2004.06.014

Lee, S. W., Shimojo, S., & O'Doherty, J. P. (2014). Neural Computations Underlying Arbitration between Model-Based and Model-free Learning. *Neuron*, *81*(3), 687–699. https://doi.org/10.1016/j.neuron.2013.11.028

Loonis, R. F., Brincat, S. L., Antzoulatos, E. G., & Miller, E. K. (2017). A Meta-Analysis Suggests Different Neural Correlates for Implicit and Explicit Learning. *Neuron*, *96*(2), 521-534.e7. https://doi.org/10.1016/j.neuron.2017.09.032

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, *24*, 167–202. https://doi.org/10.1146/annurev.neuro.24.1.167

Miyake, A., & Friedman, N. P. (2012). The Nature and Organization of Individual Differences in Executive Functions: Four General Conclusions. *Current Directions in Psychological Science*, *21*(1), 8–14. https://doi.org/10.1177/0963721411429458

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "Frontal Lobe" tasks: A latent variable analysis. *Cognitive Psychology*, *41*(1), 49–100. https://doi.org/10.1006/cogp.1999.0734

Mueller, S. T., & Piper, B. J. (2014). The Psychology Experiment Building Language (PEBL) and PEBL Test Battery. *Journal of Neuroscience Methods*, *222*, 250–259. https://doi.org/10.1016/j.jneumeth.2013.10.024

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*. https://www.frontiersin.org/articles/10.3389/fncom.2013.00147

Nemeth, D., Janacsek, K., Londe, Z., Ullman, M. T., Howard, D. V., & Howard, J. H. (2010). Sleep has no critical role in implicit motor sequence learning in young and old adults. *Experimental Brain Research*, *201*(2), 351–358. https://doi.org/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. https://doi.org/10.1093/cercor/bhs068

Osaka, M., Osaka, N., Kondo, H., Morishita, M., Fukuyama, H., Aso, T., & Shibasaki, H. (2003). The neural basis of individual differences in working memory capacity: An fMRI study. *NeuroImage*, *18*(3), 789–797. https://doi.org/10.1016/S1053-8119(02)00032-0

Otto, A. R., Gershman, S. J., Markman, A. B., & Daw, N. D. (2013). The Curse of Planning: Dissecting Multiple Reinforcement-Learning Systems by Taxing the Central Executive. *Psychological Science*, *24*(5), 751–761. https://doi.org/10.1177/0956797612463080

Otto, A. R., Raio, C. M., Chiang, A., Phelps, E. A., & Daw, N. D. (2013). Working-memory capacity protects model-based learning from stress. *Proceedings of the National Academy of Sciences*, *110*(52), 20941–20946. https://doi.org/10.1073/pnas.1312011110

Packard, M. G., Hirsh, R., & White, N. M. (1989). Differential effects of fornix and caudate nucleus lesions on two radial maze tasks: Evidence for multiple memory systems. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *9*(5), 1465–1472.

Palminteri, S., Kilford, E. J., Coricelli, G., & Blakemore, S.-J. (2016). The Computational Development of Reinforcement Learning during Adolescence. *PLOS Computational Biology*, *12*(6), e1004953. https://doi.org/10.1371/journal.pcbi.1004953

Park, J., Janacsek, K., Nemeth, D., & Jeon, H.-A. (2022). Reduced functional connectivity supports statistical learning of temporally distributed regularities. *NeuroImage*, *260*, 119459. https://doi.org/10.1016/j.neuroimage.2022.119459

Park, J., Yoon, H.-D., Yoo, T., Shin, M., & Jeon, H.-A. (2020). Potential and efficiency of statistical learning closely intertwined with individuals' executive functions: A mathematical modeling study. *Scientific Reports*, *10*(1), Article 1. https://doi.org/10.1038/s41598-020-75157-8

Parks, K. M. A., Griffith, L. A., Armstrong, N. B., & Stevenson, R. A. (2020). Statistical Learning and Social Competency: The Mediating Role of Language. *Scientific Reports*, *10*(1), Article 1. https://doi.org/10.1038/s41598-020-61047-6

Petok, J. R., Dang, L., & Hammel, B. (2022). Impaired executive functioning mediates the association between aging and deterministic sequence learning. *Aging, Neuropsychology and Cognition*, 1–17. https://doi.org/10.1080/13825585.2022.2153789

Petkov, C. I., & ten Cate, C. (2020). Structured Sequence Learning: Animal Abilities, Cognitive Operations, and Language Evolution. *Topics in Cognitive Science*, *12*(3), 828–842. https://doi.org/10.1111/tops.12444

Poldrack, R. A., Clark, J., Paré-Blagoev, E. J., Shohamy, D., Creso Moyano, J., Myers, C., & Gluck, M. A. (2001). Interactive memory systems in the human brain. *Nature*, *414*(6863), Article 6863. https://doi.org/10.1038/35107080

Poldrack, R. A., & Packard, M. G. (2003). Competition among multiple memory systems: Converging evidence from animal and human brain studies. *Neuropsychologia*, *41*(3), 245–251.

Poldrack, R. A., Prabhakaran, V., Seger, C. A., & Gabrieli, J. D. E. (1999). Striatal activation during acquisition of a cognitive skill. *Neuropsychology*, *13*, 564–574. https://doi.org/10.1037/0894-4105.13.4.564

Quentin, R., Fanuel, L., Kiss, M., Vernet, M., Vékony, T., Janacsek, K., Cohen, L. G., & Nemeth, D. (2021). Statistical learning occurs during practice while high-order rule learning during rest period. *Npj Science of Learning*, *6*(1), Article 1. https://doi.org/10.1038/s41539-021-00093-9

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

Rende, B., Ramsberger, G., & Miyake, A. (2002). Commonalities and differences in the working memory components underlying letter and category fluency tasks: A dual-task investigation. *Neuropsychology*, *16*, 309–321. https://doi.org/10.1037/0894-4105.16.3.309

Rohrmeier, M., & Rebuschat, P. (2012). Implicit Learning and Acquisition of Music. *Topics in Cognitive Science*, *4*(4), 525–553. https://doi.org/10.1111/j.1756-8765.2012.01223.x

Romano Bergstrom, J. C., Howard, J. H., & Howard, D. V. (2012). Enhanced implicit sequence learning in college-age video game players and musicians. *Applied Cognitive Psychology*, *26*(1), 91–96. https://doi.org/10.1002/acp.1800

Rosen, V. M., & Engle, R. W. (1997). The role of working memory capacity in retrieval. *Journal of Experimental Psychology: General*, *126*, 211–227. https://doi.org/10.1037/0096-3445.126.3.211

Ruffman, T., Taumoepeau, M., & Perkins, C. (2012). Statistical learning as a basis for social understanding in children. *The British Journal of Developmental Psychology*, *30*(Pt 1), 87–104. https://doi.org/10.1111/j.2044-835X.2011.02045.x

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical Learning by 8-Month-Old Infants. *Science*, *274*(5294), 1926–1928. https://doi.org/10.1126/science.274.5294.1926

Schönbrodt, F. D., & Perugini, M. (2013). At what sample size do correlations stabilize? *Journal of Research in Personality*, *47*(5), 609–612. https://doi.org/10.1016/j.jrp.2013.05.009

Schwarzer, G. (2022). Meta-Analysis in R. In *Systematic Reviews in Health Research* (pp. 510–534). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781119099369.ch26

Shao, Z., Janse, E., Visser, K., & Meyer, A. S. (2014). What do verbal fluency tasks measure? Predictors of verbal fluency performance in older adults. *Frontiers in Psychology*, *5*, 772. https://doi.org/10.3389/fpsyg.2014.00772

Shohamy, D., & Daw, N. D. (2014). Habits and reinforcement learning. In *The cognitive neurosciences, 5th ed* (pp. 591–603). MIT Press.

Shook, A., Marian, V., Bartolotti, J., & Schroeder, S. R. (2013). Musical experience influences statistical learning of a novel language. *The American Journal of Psychology*, *126*, 95–104. https://doi.org/10.5406/amerjpsyc.126.1.0095

Smalle, E. H. M., Daikoku, T., Szmalec, A., Duyck, W., & Möttönen, R. (2022). Unlocking adults' implicit statistical learning by cognitive depletion. *Proceedings of the National Academy of Sciences*, *119*(2), e2026011119. https://doi.org/10.1073/pnas.2026011119

Smalle, E. H. M., Panouilleres, M., Szmalec, A., & Möttönen, R. (2017). Language learning in the adult brain: Disrupting the dorsolateral prefrontal cortex facilitates word-form learning. *Scientific Reports*, *7*(1), Article 1. https://doi.org/10.1038/s41598-017-14547-x

Song, S., Howard, J. H., & Howard, D. V. (2007a). Implicit probabilistic sequence learning is independent of explicit awareness. *Learning & Memory*, *14*(3), 167–176. https://doi.org/10.1101/lm.437407

Song, S., Howard, J. H., & Howard, D. V. (2007b). Sleep Does Not Benefit Probabilistic Motor Sequence Learning. *Journal of Neuroscience*, *27*(46), 12475–12483. https://doi.org/10.1523/JNEUROSCI.2062-07.2007

St-Hilaire, A., Hudon, C., Vallet, G. T., Bherer, L., Lussier, M., Gagnon, J.-F., Simard, M., Gosselin, N., Escudier, F., Rouleau, I., & Macoir, J. (2016). Normative data for phonemic and semantic verbal fluency test in the adult French-Quebec population and validation study in Alzheimer's disease and depression. *The Clinical Neuropsychologist*, *30*(7), 1126–1150. https://doi.org/10.1080/13854046.2016.1195014

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

Tánczos, T., Janacsek, K., & Németh, D. (2014). [Verbal fluency tasks I. Investigation of the Hungarian version of the letter fluency task between 5 and 89 years of age]. *Psychiatria Hungarica*, *29*(2), 158–180.

The Cognitive Unconscious: The First Half Century. (2022). Oxford University Press.

Thiessen, E. D. (2011). Domain General Constraints on Statistical Learning. *Child Development*, *82*(2), 462–470. https://doi.org/10.1111/j.1467-8624.2010.01522.x

Thomson, G. H. (1939). *The factorial analysis of human ability* (pp. xv, 326). Houghton Mifflin.

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

Török, B., Nagy, D. G., Kiss, M., Janacsek, K., Németh, D., & Orbán, G. (2022). Tracking the contribution of inductive bias to individualised internal models. *PLOS Computational Biology*, *18*(6), e1010182. https://doi.org/10.1371/journal.pcbi.1010182

Tosatto, L., Fagot, J., Nemeth, D., & Rey, A. (2022). The Evolution of Chunks in Sequence Learning. *Cognitive Science*, *46*(4), e13124. https://doi.org/10.1111/cogs.13124

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. https://doi.org/10.1016/j.nlm.2017.07.015

Turk-Browne, N. B., Scholl, B. J., Johnson, M. K., & Chun, M. M. (2010). Implicit Perceptual Anticipation Triggered by Statistical Learning. *Journal of Neuroscience*, *30*(33), 11177–11187. https://doi.org/10.1523/JNEUROSCI.0858-10.2010

Ullman, M. T., Earle, F. S., Walenski, M., & Janacsek, K. (2020). The Neurocognition of Developmental Disorders of Language. *Annual Review of Psychology*, *71*(1), 389–417. https://doi.org/10.1146/annurev-psych-122216-011555

Unsworth, N., Brewer, G.A. & Spillers, G.J. Working memory capacity and retrieval from long-term memory: the role of controlled search. Mem Cogn 41, 242–254 (2013). https://doi.org/10.3758/s13421-012-0261-x

Unsworth, N., & Engle, R. W. (2005). Individual differences in working memory capacity and learning: Evidence from the serial reaction time task. *Memory & Cognition*, *33*(2), 213–220. https://doi.org/10.3758/BF03195310

Unsworth, N., Spillers, G. J., & Brewer, G. A. (2011). Variation in verbal fluency: A latent variable analysis of clustering, switching, and overall performance. *Quarterly Journal of Experimental Psychology*, *64*(3), 447–466. https://doi.org/10.1080/17470218.2010.505292

Unsworth, N., Brewer, G. A., & Spillers, G. J. (2013). Working memory capacity and retrieval from long-term memory: The role of controlled search. Memory & cognition, 41, 242-254.

Van den Bergh, D., van Doorn, J., Marsman, M., Draws, T., van Kesteren, E., Derks, K .. & Wagenmakers, E. (2020). A tutorial on conducting and interpreting a Bayesian ANOVA in JASP. *L'Année psychologique*, 120, 73-96. https://www.cairn-int.info/journal--2020-1-page-73.htm.

Vékony, T., Ambrus, G. G., Janacsek, K., & Nemeth, D. (2022). Cautious or causal? Key implicit sequence learning paradigms should not be overlooked when assessing the role of DLPFC (Commentary on Prutean et al.). *Cortex*, *148*, 222–226. https://doi.org/10.1016/j.cortex.2021.10.001

Verburgh, L., Scherder, E. J. A., van Lange, P. a. M., & Oosterlaan, J. (2016). The key to success in elite athletes? Explicit and implicit motor learning in youth elite and non-elite soccer players. *Journal of Sports Sciences*, *34*(18), 1782–1790. https://doi.org/10.1080/02640414.2015.1137344

Viechtbauer, W. (2005). Bias and Efficiency of Meta-Analytic Variance Estimators in the Random-Effects Model. *Journal of Educational and Behavioral Statistics*, *30*(3), 261–293. https://doi.org/10.3102/10769986030003261

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. https://doi.org/10.1007/s00221-015-4279-8

Voss, M. W., Prakash, R. S., Erickson, K. I., Boot, W. R., Basak, C., Neider, M. B., Simons, D. J., Fabiani, M., Gratton, G., & Kramer, A. F. (2012). Effects of training strategies implemented in a complex videogame on functional connectivity of attentional networks. *NeuroImage*, *59*(1), 138–148. https://doi.org/10.1016/j.neuroimage.2011.03.052

Wagenmakers, E.-J., Wetzels, R., Borsboom, D., & van der Maas, H. L. J. (2011). Why psychologists must change the way they analyze their data: The case of psi: Comment on Bem (2011). *Journal of Personality and Social Psychology*, *100*(3), 426–432. https://doi.org/10.1037/a0022790

West, G., Vadillo, M. A., Shanks, D. R., & Hulme, C. (2018). The procedural learning deficit hypothesis of language learning disorders: We see some problems. *Developmental Science*, *21*(2), e12552. https://doi.org/10.1111/desc.12552

Woods, S. P., Scott, J. C., Sires, D. A., Grant, I., Heaton, R. K., Tröster, A. I., & HIV Neurobehavioral Research Center Group. (2005). Action (verb) fluency: Test-retest reliability, normative standards, and construct validity. *Journal of the International Neuropsychological Society: JINS*, *11*(4), 408–415.

Xu, F., & Garcia, V. (2008). Intuitive statistics by 8-month-old infants. *Proceedings of the National Academy of Sciences*, *105*(13), 5012–5015. https://doi.org/10.1073/pnas.070445010